TRANSVERSALITY IN SPACE OF THREE DIMENSIONS*

BY

EDWARD KASNER

So-called transversality relations arose first in the calculus of variations and later in the theory of infinitesimal contact transformations. For the plane such relations are of arbitrary character, but in all spaces of more than two dimensions only correspondences of certain specific types can be identified with transversality relations. Similarly, systems of extremals, which are of arbitrary character for the simplest problem of the calculus of variations, are of peculiar geometric character for all higher problems.†

In this paper we find a simple geometric criterion for testing when a given correspondence between surface elements and line elements in three-space is of the transversality type. *Briefly stated, a certain induced homography must be involutorial. This is both necessary and sufficient.*

The result applies to simple integrals

$$\int G(x,y,z,y',z')dx = \text{minimum}, \tag{1}$$

to double integrals

$$\iint F(x,y,z,p,q)dxdy = \text{minimum}, \tag{2}$$

and to infinitesimal contact transformations, defined by a characteristic function

$$W(x,y,z,p,q). \tag{3}$$

Incidentally we obtain a *principle of transference connecting simple and double integrals in the calculus of variations*: we may associate problems (1) and (2) when they produce identical transversalities. Given the integrand function G, the other integrand function F is determined up to a factor of

---

* Presented to the Society, September 9 and October 31, 1914 (see abstracts in the Bulletin of the American Mathematical Society, vol. 21 (1914), pp. 71 and 164); received by the editors in November, 1927.

† The easiest of the higher problems, in this connection, is $\int F(x, y, y', y'')dx$ for which I have given a necessary geometric criterion in the Bulletin of the American Mathematical Society, vol. 13 (1907), pp. 289–292.

form $\mu(x, y, z)$, and vice versa. As is well known, transversality is not affected by such a factor in any of the three problems (1), (2), (3).

Contact transformations. The infinitesimal contact transformation defined by (3) carries any surface element (x, y, z, p, q) into a neighboring surface element $(x+\delta x, \cdots, q+\delta q)$, where

$$\begin{aligned} \delta x &= W_p \delta t, \\ \delta y &= W_q \delta t, \\ \delta z &= (pW_p + qW_q - W)\delta t. \end{aligned} \tag{3'}$$

If we connect the point of the old element with the point of the new element, we obtain a definite direction element (or lineal element), namely (x, y, z, y', z'), where

$$y' = \frac{W_q}{W_p}, \quad z' = \frac{pW_p + qW_q - W}{W_p}. \tag{4}$$

This lineal element is said to be transversal* to the surface element. Thus at any given point (x, y, z) we have a definite relationship or correspondence between the ∞^2 lines or directions (y', z') and the ∞^2 planes (p, q).

It is obvious that a correspondence of the peculiar analytic form (4) must have *some* geometric peculiarity, for (4) involves only one arbitrary function W of five arguments while a random correspondence would be of the general form

$$y' = \alpha(x, y, z, p, q), \quad z' = \beta(x, y, z, p, q), \tag{5}$$

involving two arbitrary functions. Before studying the appropriate geometric property, we make sure that essentially the same form (4) arises in problems (1) and (2).

Double integrals. If in the calculus of variations the boundary curve of the surface minimizing the double integral (2) is required to lie on a given surface $\phi(x, y, z) = 0$, then in the standard notation ϕ must fulfill the Kneser condition of being "transversal" to the extremal surface; and this relation

---

* The term transversal in this connection was apparently proposed first by the writer in his paper *The infinitesimal contact transformations of mechanics*, Bulletin of the American Mathematical Society, vol. 16 (1910), pp. 408–412. Vessiot, in his fundamental study of the connection between problems (1) and (3), used the term *conjugué* (see Bulletin de la Société Mathématique de France, vol. 34 (1906), pp. 230–269). Douglas presents this connection in a novel and interesting form, for both transversals and extremals, in these Transactions, vol. 29 (1927), pp. 401–420; he employs the term transversality for *any* correspondence between elements at a common point; those arising from contact transformations are described as "of special character." The relationship to double integrals was first observed by the writer.

is expressed by the fact that in every point of the boundary curve, ϕ goes through a certain direction (y', z') determined by (and duly termed transversal to) the element (p, q) of the extremal surface at that point. The formulas expressing y' and z' in terms of p and q are of exactly the same form as (1) with F in place of W.

Simple integrals. The third kind of transversality is a correspondence from line elements to surface elements. It arises from the condition which has to be fulfilled at the end point of the curve minimizing the integral (1) if this end point is required to lie on a given surface $\psi(x, y, z) = 0$. The transversality of the element (p, q) of ψ to the direction (y', z') of the curve at that point is expressed by the relations

$$p = \frac{y'G_{y'} + z'G_{z'} - G}{G_{z'}}, \quad q = -\frac{G_{y'}}{G_{z'}}. \tag{6}$$

By means of the simple transformation

$$\begin{aligned} P &= -z', \quad Q = y', \\ Y' &= q, \qquad Z' = -p, \\ \overline{G}(P,Q) &= G(Q, -P), \end{aligned} \tag{7}$$

we see that (6) is reduced to a correspondence of the form (4) from the surface elements (P, Q) to the line elements (Y', Z'), with the new function $\overline{G}$ taking the place of W.

Thus the three kinds of transversality essentially reduce to one, expressed by the form (4) *. To every function $W(x, y, z, p, q)$ corresponds a definite transversality, but conversely a given transversality determines W only up to an arbitrary function of x, y, z as a factor.*

The geometric criterion. The peculiar character of (4) admits a simple geometric interpretation. Let us take a general analytic correspondence (5) associating surface and line elements. Denote it by T.

Consider an arbitrary point and through it a plane determined by (p, q) as fixed. By means of T a certain *induced* correspondence H arises in the pencil of directions determined by the point and the plane. It associates with the intersection of (p, q) and the neighboring element $(p+dp, q+dq)$ the intersection of (p, q) with the plane determined by the two directions (y', z') and $(y'+dy', z'+dz')$ corresponding to (p, q) and $(p+dp, q+dq)$ by T. This correspondence is easily seen to be a homography for every transformation T. When (5) is of the special type (4) we shall show that *this homography is involutorial*, and that the converse is also valid.

Transversality is therefore distinguished geometrically among all transformations (5) *by the property that the induced homography H arising in the above described manner in every pencil of directions is an involution.*

The calculations leading to the characteristic involution are as follows. We first discuss the equations

$$\alpha = \frac{W_q}{W_p}, \quad \beta = \frac{pW_{p+q}W_q - W}{W_p},$$

obtained by comparing (4) and (5). The easiest way to eliminate W is to form

$$\frac{W_p}{W} = \frac{1}{p + q\alpha - \beta}, \quad \frac{W_q}{W} = \frac{\alpha}{p + q\alpha - \beta},$$

and then write the condition for integrability

$$\frac{\partial}{\partial q}\left(\frac{1}{p + q\alpha - \beta}\right) - \frac{\partial}{\partial p}\left(\frac{\alpha}{p + q\alpha - \beta}\right) = 0.$$

We obtain

$$\alpha + q\alpha_q - \beta_q + \alpha_p(p + q\alpha - \beta) - \alpha(1 + q\alpha_p - \beta_p) = 0,$$

or, finally,

$$(p - \beta)\alpha_p + q\alpha_q + \alpha\beta_p - \beta_q = 0, \tag{8}$$

which is therefore the criterion for transversality in analytic form.

The induced homography H. Consider an arbitrary correspondence

$$y' = \alpha(p,q), \quad z' = \beta(p,q) \tag{9}$$

between the surface elements and direction elements at a given point. (Since the point is fixed it will not be necessary to write the x, y, z arguments in α and β.) To the element (p, q) whose plane is

$$p\delta x + q\delta y - \delta z = 0$$

corresponds the direction

$$1 : \alpha : \beta ;$$

to a neighboring element $(p+dp, q+dq)$, corresponds the direction

$$1 : \alpha + d\alpha : \beta + d\beta,$$

where $d\alpha = \alpha_p dp + \alpha_q dq$, $d\beta = \beta_p dp + \beta_q dq$. The plane of the two direction elements is

$$(\alpha d\beta - \beta d\alpha)\delta x - d\beta\delta y + d\alpha\delta z = 0.$$

It cuts the plane $p\delta x+q\delta y-\delta z=0$ in the direction

$$(m) \qquad qd\alpha - d\beta : (\beta - p)d\alpha - \alpha d\beta : q\beta d\alpha - (p + q\alpha)d\beta.$$

On the other hand $p\delta x+q\delta y-\delta z=0$ cuts the neighboring plane $(p+dp)\delta x+(q+dq)\delta y-\delta z=0$ in

$$(M) \qquad dq : -dp : pdq - qdp.$$

Thus we have in the plane (p, q) supposed fixed a definite induced correspondence H from the direction (m) to the direction (M). This H is obviously a homography since the differentials dp, dq enter linearly and p and q are fixed. The result holds for all analytic transformations (5) or (9).

To prove that for the special correspondence (4) this homography H is involutorial, it is sufficient to express one of the ratios of m in terms of the corresponding ratio of M. Putting

$$m_1 = \frac{(p-\beta)d\alpha + \alpha d\beta}{qd\alpha - d\beta} = \frac{\{(p-\beta)\alpha_p + \alpha\beta_p\}dp + \{(p-\beta)\alpha_q + \alpha\beta_q\}dq}{(q\alpha_p - \beta_p)dp + (q\alpha_q - \beta_q)dq}$$

and $M_1=dp/dq$ we have

$$m_1 = \frac{\{(p-\beta)\alpha_p + \alpha\beta_p\}M_1 + \{(p-\beta)\alpha_q + \alpha\beta_q\}}{(q\alpha_p - \beta_p)M_1 + (q\alpha_q - \beta_q)}.$$

The condition that this shall be an involution is

$$(p-\beta)\alpha_p+\alpha\beta_p+q\alpha_q-\beta_q=0.$$

This is identical with the condition for transversality obtained in (8). Therefore our theorem is proved.

For the general analytic transformation

$$p = \alpha(x,y,z,y',z'), \quad q = \beta(x,y,z,y',z')$$

from a line element to a surface element, we find of course the corresponding dual result: it determines a homography in the pencil of planes through the direction (y', z') supposed fixed, and this homography is an involution when and only when the correspondence is a transversality.

Linear transversalities and geodesic systems. If a correspondence from plane elements to line elements is assumed to be projective, when will it satisfy the condition for transversality? Using the analytic test (8) we readily find that the correspondence is duality (with respect to any quadric cone at the given point). Of course this is not true for two dimensions, but it is true for all higher spaces.

The only case in which space transversality is a projective relation is where it reduces to duality.

The integrand functions G and F in (1) and (2), and the characteristic function (W) in (3) then reduce to the special form of the square root of any quadratic polynomial (so that the extremals are the geodesics of a Riemann manifold). This includes the more special case of natural families of trajectories,

$$\int \nu(x,y,z)(1 + y'^2 + z'^2)^{1/2}dx = \text{minimum},$$

where transversality is ordinary orthogonality.

I showed in 1910 that the ∞^4 extremals are then distinguished by the fact that the ∞^2 which are orthogonal to any base surface always admit ∞^1 orthogonal surfaces.* This has been proved more briefly by Schouten,† who uses the term conform-geodetic instead of my term natural. Douglas (analytically) and Blaschke (synthetically)‡ have recently shown that such a Kneser integrability test is valid in the characterization of any extremal family of curves (problem (1) above), orthogonality being of course replaced by transversality.

It follows from this theorem and the theorem above on polarities, that *any geodesic system (that is, any family of curves in flat space obtainable by arbitrary point representation of the geodesics of a Riemann space) may be characterized by the existence at each point of a projective (necessarily polar) transversality with respect to which the system has the integrability property.*

Synthetic treatment. I observe in conclusion that the geometric criterion of general transversality (involutorial character of H) may be deduced synthetically from another interpretation of the analytic test (8). If a line element at a given point corresponds to a surface element, and if we construct parallel surface elements at all the ∞^1 points of the straight line determined by the line element, we obtain in all ∞^3 surface elements, that is, a Pfaffian equation; the integrability of this equation (for every given point) is the necessary and sufficient criterion of transversality. (See also the equivalent statement quoted by Douglas, page 403.) The ∞^1 integral surfaces are of course parallel, and we may use the theory of conjugate directions to obtain the property above, which has the advantage of dealing only with the elements at a given point.

* This is the converse of the Thomson-Tait theorem, discussed in *the Theorem of Thomson and Tait and natural families of trajectories*, these Transactions, vol. 11 (1910), pp.121–140.

† Nieuw Archief voor Wiskunde, 1927.

‡ Douglas in the paper cited above, p. 404; Blaschke, *Ueber die Umkehrung der Kneser'schen Satz*, Jahresbericht der Deutschen Mathematiker-Vereinigung, 1927.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ON INTEGRAL EQUATIONS WITH DISCONTINUOUS KERNELS*

BY

J. D. TAMARKIN AND RUDOLPH E. LANGER

1. **Introduction.** In recent publications the authors have individually considered boundary value and expansion problems arising from the integral equation of the type

$$y(t) = \rho \int_{\alpha}^{\beta} \Gamma(t,\tau)y(\tau)d\tau, \tag{1}$$

when the kernel presents certain features of finite discontinuity along the line $t=\tau$. Kernels of this kind are typified by the Green's functions of ordinary differential equations with boundary conditions, and it was first noticed by Birkhoff that this fact might serve as a basis for the development of a theory of integral equations of this type. The discussion of the case in which the kernel itself is discontinuous was given by Langer† and more recently the discussion of the case in which the discontinuities occur in the first partial derivatives of the kernel, the kernel itself being continuous, was given by Langer jointly with Mrs. E. P. Brown.‡ A history of the problem may be found in the introductions to these papers. The method of attack in each case depends upon a transformation of the integral equation into an integro-differential system.

A theory of integro-differential systems composed of an equation with boundary conditions has meanwhile been given independently by Tamarkin,§ who showed also the applicability of this theory to the integral equation in which the kernel is discontinuous by giving a reduction (distinct from that of the papers mentioned above) of the integral equation to a system of the type considered.

The present paper is given to the discussion of the equation of type (1) in which the kernel and its partial derivatives of order $(n-2)$ are continuous

---

* Presented to the Society, December 28, 1926; received by the editors in December, 1927.

† *On the theory of integral equations with discontinuous kernels*, these Transactions, vol. 28 (1926), pp. 585–639.

‡ *On a class of integral equations with discontinuous kernels*, these Transactions, vol. 29 (1927), pp. 683–715.

§ *The notion of the Green's function in the theory of integro-differential equations*, these Transactions, vol. 29 (1927), pp. 755–800.

and the discontinuities along the line $t=\tau$ occur in the partial derivatives of higher order. Of the papers mentioned above those by Tamarkin and Langer respectively will be referred to in the text by [T] and by [L]. By [B] we refer to a paper by Birkhoff* of which we make frequent use.

2. **The given equation.** We consider the integral equation (1) supposing the kernel to be possessed of the following properties in the region

$$\alpha \leqq \begin{matrix} t \\ \tau \end{matrix} \leqq \beta.$$

(i) The kernel and its partial derivatives to those of order $(n-2)$ $(n \geqq 2)$ are continuous in (t, τ).

(ii) The partial derivatives of orders $(n-1)$ to $3n$ exist, are continuous except on the line $t=\tau$, and as the point (t, τ) approaches any point (t_0, t_0) along a path having no points in common with the line $t=\tau$ these derivatives tend to definite limiting values which are independent of the path and are continuous in t.

(iii) $$\frac{\partial^{n-1}\Gamma(t,\tau)}{\partial t^{n-1}}\Bigg]_{\tau=t-}^{\tau=t+} \equiv \phi(t) \neq 0.$$

Under these hypotheses the function $\phi(t)$ is continuous and we may assume without loss of generality that

$$\phi(t) \geqq \phi > 0, \tag{2}$$

since this involves at most a change in the sign of ρ.

3. **The normalization of the equation.** Since $\phi(t)$ satisfies the relation (2) its integral is monotonic increasing, and hence, the independent variable of the given equation may be changed by the substitution [L, §3]

$$x = \frac{1}{l}\int_\alpha^t \phi(t)^{1/n}dt, \quad \xi = \frac{1}{l}\int_\alpha^\tau \phi(\tau)^{1/n}d\tau, \tag{3}$$

where

$$l = \int_\alpha^\beta \phi(t)^{1/n}dt.$$

This reduces the equation to the form

$$u(x) = \lambda \int_0^1 K(x,\xi)u(\xi)d\xi, \tag{4}$$

---

* *Boundary value and expansion problems* etc., these Transactions, vol. 9 (1908), pp. 373–395.

where we have set

$$K(x,\xi) \equiv l^{-n+1}\Gamma(t,\tau)\phi(\tau)^{-1/n},$$
$$\lambda = l^n\rho,$$
$$u(x) \equiv y(t), \quad u(\xi) \equiv y(\tau).$$

We shall designate the equation in the form (4) as the "normalized equation." The properties of its kernel are readily found to be the following:

(A) The kernel and its partial derivatives to those of order $(n-2)$ are continuous in (x, ξ).

(B) The partial derivatives of orders $(n-1)$ to $(2n+1)$ exist and are continuous except for $x=\xi$, and as the point (x, ξ) approaches any point (x_0, x_0) along a path having no points in common with the line $x=\xi$ these derivatives tend to definite limiting values which are independent of the path and are continuous in x.

(C) If we abbreviate by setting

$$J_{i,j}f \equiv D_x^i D_\xi^j f(x,\xi)\Big]_{\xi=x-}^{\xi=x+}$$

then

$$J_{n-1-s,s}K \equiv (-1)^s \qquad (s = 0,1,2,\cdots,n-1).$$

The proof of (C) is easily made to depend on the following lemma [L, 2].*

LEMMA. *If $F(t, \tau)$ is any function whose partial derivatives of the first order are possessed of property* (ii) *above, then*

$$\frac{dF(t,t\pm)}{dt} = F_t(t,t\pm) + F_\tau(t,t\pm).$$

4. **Transformation of the integral equation into an integro-differential equation.**† In virtue of the properties of the kernel $K(x, \xi)$ we find by differentiation of equation (4) that

$$u^{(s)}(x) = \int_0^1 D_x^s K(x,\xi)\lambda u(\xi)d\xi, \qquad \text{for } s = 0,1,2,\cdots,n-1, \tag{5}$$

$$u^{(n)}(x) = -\lambda u(x) + \int_0^1 D_x^n K(x,\xi)\lambda u(\xi)d\xi. \tag{6}$$

---

* Cf. also Hammerstein, *Über die asymptotische Darstellung der Eigenfunktionen linearer Integralgleichungen*, Mathematische Annalen, vol. 93 (1924), p. 113.

† The method is that of [T, §§29–31]. Cf. also the method of achieving such a transformation in [L, Ch. 3] and Langer and Brown, loc. cit., Ch. 3.

We shall make now the further hypothesis

(iv) The given equation is such that the kernel

$$E(x,\xi) \equiv D_x^n K(x,\xi)$$

is possessed of a reciprocal kernel $\mathfrak{E}(x, \xi)$, defined by

$$E(x,\xi) + \mathfrak{E}(x,\xi) = \int_0^1 E(x,s)\mathfrak{E}(s,\xi)ds = \int_0^1 \mathfrak{E}(x,s)E(s,\xi)ds.$$

Then considering (6) formally as an equation in $\lambda u(x)$ with $u^{(n)}(x)$ known, we may write it in the "solved" form

$$\lambda u(x) = -u^{(n)}(x) + \int_0^1 \mathfrak{E}(x,\xi)u^{(n)}(\xi)d\xi. \tag{7}$$

It is readily verified that $\mathfrak{E}(x, \xi)$ possesses the property (B) of the normalized kernel with $(2n+1)$ replaced by $(n+1)$. The repeated integration by parts of the integral in (7) yields for that equation the form

$$L(u) + \lambda u = \sum_{i=1}^{n} \{u^{(i-1)}(0)\phi_{0,i}(x) + u^{(i-1)}(1)\phi_{1,i}(x)\} + \int_0^1 f(x,\xi)u(\xi)d\xi, \tag{8}$$

where

$$L(u) = u^{(n)}(x) + p_1(x)u^{(n-1)}(x) + \cdots + p_n(x)u(x),$$

$$p_i(x) = (-1)^{i-1}J_{0,i-1}\mathfrak{E},$$

$$\phi_{0,i}(x) = -(-1)^{n-i}D_\xi^{n-i}\mathfrak{E}(x,0),\ \phi_{1,i}(x) = (-1)^{n-i}D_\xi^{n-i}\mathfrak{E}(x,1),$$

$$f(x,\xi) = (-1)^n D_\xi^n \mathfrak{E}(x,\xi).$$

5. **The boundary conditions.** Every solution of (4) thus satisfies also the equation (8). In addition, however, it satisfies certain boundary conditions which may be obtained as follows.* Set $x=0$ in equations (5) and substitute in the integral the value of $\lambda u(\xi)$ as given formally by solving (8) for the second term on its left. With the use of Green's identity

$$\int_0^1 \{vL(u) - uL'(v)\}dx = P(u,v),$$

where L' is the operator adjoint to L, and $P(u, v)$ is the corresponding bilinear form in $u^{(j)}(0)$, $u^{(j)}(1)$ and $v^{(j)}(0)$, $v^{(j)}(1)$, $j=1, 2, \cdots, (n-1)$, we readily obtain the result in the form

---

* L, §§ 10, 31; T, § 29.

$$L_i(u) = \int_0^1 \alpha_i(\xi)u(\xi)d\xi \qquad (i = 1,2,\cdots,n), \tag{9}$$

where

$$L_i(u) = u^{(i-1)}(0) + P(u,K_i) - \sum_{j=1}^{n}\Big\{u^{(j-1)}(0)\int_0^1 K_i(\xi)\phi_{0,j}(\xi)d\xi + u^{(j-1)}(1)\int_0^1 K_i(\xi)\phi_{1,j}(\xi)d\xi\Big\},$$

$$\alpha_i(\xi) = -L'(K_i) + \int_0^1 K_i(t)f(t,\xi)dt, \tag{10}$$

$$K_i(\xi) \equiv D_x^{i-1}K(0,\xi).$$

We state this result as follows.

THEOREM 1. *Every solution of the normalized integral equation* (4) *is also a solution of the integro-differential system composed of the equation* (8) *and the boundary conditions* (9).

6. **The equivalence of the integro-differential system and the integral equation.** We consider now the non-homogeneous system

$$L(u) + \lambda u = f(x) + \sum_{j=1}^{n}\{u^{(j-1)}(0)\phi_{0,j}(x) + u^{(j-1)}(1)\phi_{1,j}(x)\} + \int_0^1 f(x,\xi)u(\xi)d\xi, \tag{11}$$

together with the boundary conditions (9). The equation (11) differs from (8) only in the presence of the term $f(x)$ which is taken to be an arbitrary continuous function. We shall show that this system is equivalent to a non-homogeneous integral equation of the second kind.

Since (8) and (7) are equivalent forms of the same equality it is clear that (11) is equivalent to the equation

$$u^{(n)}(x) = f(x) - \lambda u(x) + \int_0^1 \mathfrak{E}(x,\xi)u^{(n)}(\xi)d\xi,$$

and since $E(x, \xi)$ is the resolvent kernel of $\mathfrak{E}(x, \xi)$ this equation may be "solved" to give

$$u^{(n)}(x) = f(x) - \lambda u(x) - \int_0^1 E(x,\xi)\{f(\xi) - \lambda u(\xi)\}d\xi. \tag{12}$$

Recalling that (6) is obtained from (4) by an n-fold differentiation it becomes clear that (12) may be written

$$u^{(n)}(x) = \frac{d^n}{dx^n}\int_0^1 K(x,\xi)\{\lambda u(\xi) - f(\xi)\}d\xi. \tag{13}$$

Let the values (10) be substituted now into the equations (9). These become then

$$u^{(i)}(0) = \int_0^1 D_x^i K(0,\xi)\Big\{ - L(u) + \sum_{j=1}^{n}[u^{(j-1)}(0)\phi_{0,j}(\xi) + u^{(j-1)}(1)\phi_{1,j}(\xi)] + \int_0^1 f(\xi,t)u(t)dt\Big\}d\xi,$$

or, in virtue of (11),

$$u^{(i)}(0) = \int_0^1 D_x^i K(0,\xi)\{\lambda u(\xi) - f(\xi)\}d\xi \qquad (i = 0,1,2,\cdots,n-1).$$

With this fact it becomes clear that in each of the first n successive integrations of (13) the constant of integration is zero, so that we obtain from (13)

$$u(x) = F(x) + \lambda\int_0^1 K(x,\xi)u(\xi)d\xi, \tag{14}$$

where

$$F(x) = -\int_0^1 K(x,\xi)f(\xi)d\xi.$$

Hence every solution of the system (11), (9) is also a solution of equation (14).

Lastly it is clear that the system composed of (11) and (9) is not intrinsically altered if we replace in (11) every quantity of the set $u^{(j)}(0)$, $u^{(j)}(1)$, $j=0,1,2,\cdots,(n-1)$, which is expressible linearly in terms of the quantities $L_i(u)$, by their values as given by (9). If this is done, the system (11), (9) may be written in the form

$$\text{(a)}\quad L(u) + \lambda u = f(x) + \sum_{j=0}^{m} l_j(u)\Phi_j(x) + \int_0^1 h(x,\xi)u(\xi)d\xi,$$

(15)

$$\text{(b)}\quad L_i(u) = \int_0^1 \alpha_i(\xi)u(\xi)d\xi \qquad (i = 1,2,\cdots,n),$$

where the $l_j(u)$, $j=1, 2, \cdots, m$, are simply those of quantities $u^{(k)}(0)$, $u^{(k)}(1)$ which are linearly independent of the quantities $L_i(u)$.

We have shown, then, that every solution of (15) is a solution of (14). The converse is easily shown by treating equation (14) as the corresponding

homogeneous equation (4) was treated in §4. Hence we may state the following theorem:

THEOREM 2. *Under the hypothesis* (iv) *the integral equation* (14) *and the integro-differential system* (15) *are equivalent in the sense that every solution of either is also a solution of the other.*

COROLLARY. *Under the hypothesis* (iv) *the given integral equation* (4) *is equivalent to the homogeneous differential system obtained from* (15) *by setting* $f(x) \equiv 0$.

7. **The Green's function of system (15).** The integro-differential system of a type under which system (15) is included has been studied in paper [T]. To ascertain the availability of the results obtained there we must consider the hypotheses made [T: §10, A, i–v; §13, B, i–ii; §19, C, i–v] and whether they are satisfied by the system in hand. Due to the specialized structure of system (15) it is readily found that of these hypotheses some [T: A, i, iii; B, i; C, i–iii, v] are automatically satisfied. To meet the conditions [T: A, ii, v; C, iv] we must make the following hypothesis:

(v) The partial derivatives of order $(2n+1)$ of $K(x, \xi)$ are of bounded variation in ξ for fixed x and in x for fixed ξ, and are of uniformly bounded total variation on $(0, 1)$.

To meet the remaining conditions [T: A, iv; B, ii] we must suppose further that the given integral equation is such that

(vi) The auxiliary differential system

$$L(u) + \lambda u = f(x), \qquad L_i(u) = 0 \qquad (i = 1, 2, \cdots, n) \tag{15'}$$

is regular in the sense of Birkhoff [B, pp. 382–383].

By Theorem 2 the system (15) and the equation (14) have the same characteristic values. Let us suppose that λ is not such a value. Then the system (15) admits of a unique solution [T, Theorems 1, 2, 3] which is expressible in the form

$$u(x) = \int_0^1 \Gamma(x, t, \lambda) f(t) dt, \tag{16}$$

where $\Gamma(x, t, \lambda)$ is the Green's function of system (15).* On the other hand since λ is not a characteristic value the equation (14) possesses a resolvent kernel $\mathfrak{K}(x, t, \lambda)$ in terms of which the equation may be written

* The notation here differs slightly from that in [T, Ch. 1] where the Green's function is written $\Gamma(x, t, \rho)$ while $\rho^n = \lambda$.

$$u(x) = F(x) - \lambda \int_0^1 \mathfrak{K}(x,t,\lambda)F(t)dt.$$

Substituting in this the value of $F(x)$, and recalling the relation

$$K(x,t) + \mathfrak{K}(x,t,\lambda) = \lambda \int_0^1 \mathfrak{K}(x,\xi,\lambda)K(\xi,t)d\xi,$$

we find that

$$u(x) = \int_0^1 \mathfrak{K}(x,t,\lambda)f(t)dt. \tag{17}$$

Now $\Gamma(x, t, \lambda)$ and $\mathfrak{K}(x, t, \lambda)$ are both meromorphic functions of λ with poles at the characteristic values of system (15). From this fact together with the fact that the right hand members of (16) and (17) are identical for an arbitrary choice $f(t)$, it follows that

$$\Gamma(x,t,\lambda) \equiv \mathfrak{K}(x,t,\lambda).$$

8. **The characteristic values.** We consider now in conjunction with the system (15) the associated differential system (15′). If we denote the Green's function of (15′) by $\Gamma'(x, t, \lambda)$ we find directly that (15′) is expressible in the form of the integral equation

$$u(x) = F'(x) + \lambda \int_0^1 K'(x,t)u(t)dt, \tag{14′}$$

where

$$K'(x,t) \equiv -\Gamma'(x,t,0),$$

and

$$F'(x) \equiv -\int_0^1 K'(x,t)f(t)dt.$$

There is no loss in generality in supposing that $\lambda=0$ is not a characteristic value of system (15′), since a translation of the plane would not affect the character of the system.

It is clear that the resolvent kernel of $K'(x, t)$ is $\Gamma'(x, t, \lambda)$. Hence if $\lambda=\lambda_r'$ is any characteristic value of the equation (14′) we draw from the theory of the Fredholm equation* the fact that the coefficient of $(\lambda-\lambda_r')^{-1}$ in the expansion of $\Gamma'(x, t, \lambda)$ in ascending powers of $(\lambda-\lambda_r')$ is

$$\sum_{i=1}^{\sigma_r'} \phi_i(x)\psi_i(t), \tag{18}$$

where $\{\phi_i, \psi_i\}$ is the biorthogonal set of fundamental functions of (14′)

* Cf. Lalesco, *Introduction à la Théorie des Équations Intégrales*, Paris, 1912, pp. 40–59.

corresponding to $\lambda=\lambda_\nu'$.* These functions are characteristic functions, i.e. solutions of the equation, when and only when $\lambda=\lambda_\nu'$ is a simple pole of $\Gamma'(x, t, \lambda)$. The integer σ_ν' is called the multiplicity of the corresponding characteristic value.

Let the subscripts be assigned to the characteristic values now so that

$$|\lambda_1'| \leqq |\lambda_2'| \leqq \cdots, \tag{19'}$$

each value λ_ν' being repeated in this array a number of times equal to its index of multiplicity. The corresponding biorthogonal system of fundamental functions we designate by

$$\{u_\nu'(x), v_\nu'(x)\} \qquad (\nu = 1,2, \cdots).$$

In similar fashion we arrange the characteristic values and fundamental functions of equation (14) in the array

$$\begin{gathered} |\lambda_1| \leqq |\lambda_2| \leqq \cdots, \\ \{u_\nu(x), v_\nu(x)\}. \end{gathered} \tag{19}$$

Let the positive constant δ be chosen now arbitrarily small and let the circles c_ν' with radii δ and centers at the points λ_ν' be drawn in the λ plane. It is known [T, §23] that there exist infinitely many characteristic values λ_ν and moreover that all those characteristic values which exceed in absolute value a certain constant Λ (which depends on δ) lie within the circles c_ν'.

Due to the similarities of structure of the systems (15) and (15′) the statement above may be made somewhat more precise. The set of values (17′) are given by formulas [B, p. 383-385]

$$\lambda_\nu' = -(\pm 2\nu\pi i)^n\left(1+\frac{\alpha}{\nu}+\frac{E}{\nu^2}\right), \tag{20}$$

where α is a constant. Hence the constant δ may be chosen so small that for $|\lambda_\nu'|$ sufficiently large no two of the small circles c_ν' have points in common. Then since the residue of $\Gamma'(x, t, \lambda)$ at λ_ν' is given by (18), the set $\{\phi_i, \psi_i\}$ being biorthogonal we have

$$\frac{1}{2\pi i}\int_{c_\nu'}\int_0^1 \Gamma'(x,x,\lambda)dxd\lambda = \sigma_\nu'. \tag{21'}$$

Similarly we have

$$\frac{1}{2\pi i}\int_{c_\nu'}\int_0^1 \Gamma(x,x,\lambda)dxd\lambda = \sigma_\nu, \tag{21}$$

* These functions are called principal functions in the discussion of a more general boundary problem not linear in λ. Cf. Tamarkin, *Some general problems of the theory of ordinary linear differential equations*, etc., Mathematische Zeitschrift, vol. 27 (1927), pp. 1–54.

and σ_ν represents the number of members of the set (19) which lie within the circle c_ν'; σ_ν' represents, of course, the number of members in set (17') which lie in c_ν'.

On the other hand if we set $\rho^n=\lambda$ and consider for the moment the situation in the ρ-plane, it is known [T, §§19–21] that the dominant terms of $\Gamma'(x, t, \lambda)$ and $\Gamma(x, t, \lambda)$ are the same (they depend only on the differential operator $L(u)$) provided that $|\rho|$ is sufficiently large and that $|\rho-\bar{\rho}|$ is uniformly bounded from zero, where $\bar{\rho}$ is any characteristic value ρ_ν' or ρ_ν. Moreover, in this case the difference of these Green's functions is of the order of ρ^{-n}. Hence we may write, when λ is a point of a circle c_ν', and $|\lambda|$ is sufficiently large,

$$|\Gamma'(x,t,\lambda) - \Gamma(x,t,\lambda)| < N/|\lambda|,$$

where N is a constant. But from this and (21), (21') it follows that

$$|\sigma_\nu' - \sigma_\nu| \leqq N/r_\nu$$

where r_ν is the minimum of $|\lambda|$ on the circle c_ν'. Since the numbers σ_ν' and σ_ν are integers whereas r_ν increases indefinitely with ν, it follows that for ν sufficiently large $\sigma_\nu'=\sigma_\nu$, i.e. the number of values λ_j counted with their multiplicities which lie in c_ν' is equal to the multiplicity of λ_ν' when $|\lambda_\nu'|>\Lambda$.

As to the actual multiplicity of the value λ_ν' we note that because of the hypothesis (vi) these values are in general simple and can be at most double. A necessary though not sufficient condition for double characteristic values is that n be even and that the values θ_i defined by Birkhoff [B, p. 383]* satisfy the relation

$$\theta_1^2 - 4\theta_0\theta_2 = 0.$$

We formulate these results as follows.

THEOREM 3. *If the kernel of the normalized integral equation* (4) *satisfies the hypotheses* (iv)-(vi), *then there exist infinitely many characteristic values* $\lambda_j, j=1, 2, \cdots$. *These values are represented asymptotically by the quantities* λ_ν' *given by* (20) *in the sense that for* $|\lambda|$ *sufficiently large all values* λ_j *lie within circles* c_ν' *of arbitrarily small radii* δ *about the quantities* λ_ν' *as centers. The characteristic values are in general simple and in general but one lies in a circle* c_ν'. *A necessary condition for double characteristic values or that circles* c_ν' *enclose two characteristic values is that* n *be even and that*

$$\theta_1^2 - 4\theta_0\theta_2 = 0.$$

* Cf. also Tamarkin, *Sur quelques points de la théorie des équations différentielles* etc., Rendiconti del Circolo Matematico di Palermo, vol. 23 (1912), pp. 345–382.

9. **The characteristic functions.** In the deduction of the asymptotic expressions for the characteristic functions the formal work for the case when the integer n is odd differs to some extent from that for the case when n is even. To avoid repetition we shall treat only the case in which n is odd. The reader will find no difficulty with the parallel deductions for the omitted case.

By the corollary to Theorem 2 the given integral equation is equivalent to the homogeneous system (15). Let

$$y_1(x), y_2(x), \cdots, y_n(x)$$

be any fundamental set of solutions of the differential equation

$$L(y) + \lambda y = 0. \tag{22}$$

Further let $\delta(t)$ designate the determinant

$$\delta(t) \equiv \begin{vmatrix} y_1^{(n-1)}(t) & \cdots & y_n^{(n-1)}(t) \\ \vdots & & \vdots \\ y_1(t) & \cdots & y_n(t) \end{vmatrix},$$

and $\delta_j(t)$ the cofactor in $\delta(t)$ of the element $y_j^{(n-1)}(t)$. Then upon setting

$$z_k(t) = \delta_k(t)/\delta(t) \qquad (k = 1, 2, \cdots, n),$$

the general solution of the non-homogeneous equation

$$L(u) + \lambda u = \phi(x) \tag{23}$$

may be written in the form

$$u(x) = \sum_{j=1}^{n} c_j y_j(x) + \int_0^1 \gamma(x,t)\phi(t)dt, \tag{24}$$

where $c_j, j = 1, 2, \cdots, n$, are arbitrary constants, and

$$\gamma(x,t) = \begin{cases} \sum_{j=1}^{\mu} y_j(x)z_j(t) & \text{for } t < x, \\ \sum_{j=\mu+1}^{n} y_j(x)z_j(t) & \text{for } t > x. \end{cases} \tag{25}$$

In this μ may be chosen as any integer of the set $1, 2, \cdots, n$.

Let λ be chosen now as a characteristic value of equation (4) and set $\rho^n = \lambda$. Further let $\omega_j, j = 1, 2, \cdots, n$, designate the n roots of the equation $\omega^n + 1 = 0$, the subscripts being so assigned that

$$R(\rho\omega_1) \leqq R(\rho\omega_2) \leqq \cdots \leqq R(\rho\omega_n).$$

The integer μ in (25) we shall choose so that

$$R(\rho\omega_{\mu-1}) < 0 < R(\rho\omega_{\mu+1}).$$

There exists then [B, p. 391] a fundamental set of solutions $y_j(x)$ of equation (22) whose asymptotic form is given by the relations

$$\frac{d^s y_j(x)}{dx^s} = e^{\rho\omega_j x}\rho^s[\omega_j{}^s], \quad *$$

$$\frac{d^s z_j(t)}{dt^s} = \frac{1}{n}e^{-\rho\omega_j t}\rho^{s-n+1}[(-\omega_j)^{s+1}] \qquad (s = 0,1,\cdots,n-1). \tag{26}$$

We define the quantities $\tau_s(x,t)$ now by the formulas

$$\tau_s(x,t) = \begin{cases} \sum_{j=1}^{\mu} e^{\rho\omega_j(x-t)}\left[\frac{-\omega_j{}^s}{n}\right], & \text{when } t < x, \\ \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-t)}\left[\frac{\omega_j{}^s}{n}\right], & \text{when } t > x. \end{cases} \tag{27}$$

Then we obtain from (25) and (26) the relation

$$\gamma(x,t) = \rho^{-n+1}\tau_1(x,t). \tag{28}$$

Now by Theorem 3, $|\rho-\rho_\nu'|$ is uniformly small when $|\rho|$ is sufficiently large, $\rho_\nu'^n=\lambda_\nu'$ being a characteristic value of the system (15′). With the use of formula [B, p. 385]

$$\rho_\nu' = \frac{2\nu\pi i}{\omega_\mu}\left\{1+\frac{a}{\nu}+\frac{E}{\nu^2}\right\}, \tag{29}$$

a being a constant, it is readily found therefore that the quantities

$$e^{\rho\omega_j(x-t)}, \qquad \begin{matrix} t < x \text{ for } j = 1,2,\cdots,\mu, \\ t > x \text{ for } j = \mu+1,\cdots,n, \end{matrix}$$

are uniformly bounded.

Returning to equation (23) let $\phi(x)$ be chosen as the function

$$\text{(a)} \quad \phi(x) = \sum_{j=1}^{m} c_j'\,\Phi_j(x) + \int_0^1 h(x,\xi)u(\xi)d\xi,$$

$$\text{(b)} \quad c_j' = l_j(u), \qquad (j = 1,2,\cdots,m), \tag{30}$$

* We use here the notation of [B, p. 389] wherein $[\omega]$ designates an expression of the form $\omega+E'\rho$ and E is bounded for $|\rho|$ large.

the quantities involved being those which occur also in equation (15a). The equation (23) becomes then identically the equation (15a) with $f(x) \equiv 0$. Hence it is found on substituting the forms (25) in equation (24) that the solution of the homogeneous equation (15a) satisfies the relation

$$u(x) = \sum_{j=1}^{n} c_j y_j(x) + \sum_{j=1}^{m} c_j' \Psi_j(x) + \int_0^1 H(x,\xi) u(\xi) d\xi, \tag{31}$$

where

$$\Psi_j(x) = \int_0^1 \gamma(x,t) \Phi_j(t) dt,$$
$$H(x,\xi) = \int_0^1 \gamma(x,t) h(t,\xi) dt. \tag{32}$$

Under the hypothesis (v) the functions $\Phi_j(t)$ are continuous and have derivatives of bounded variation on (0, 1), while $h(t, \xi)$ as a function of t with ξ fixed has the same properties on each of the sub-intervals $(0, \xi)$ and $(\xi, 1)$.

Consider the integral $H(x, \xi)$. Integrating the formula (32) by parts we obtain

$$\begin{aligned} H(x,\xi) = -\rho^{-n}\{ &- h(x,\xi)[1] + \tau_0(x,\xi)[h(\xi-,\xi) - h(\xi+,\xi)] \\ &- \tau_0(x,0)h(0,\xi) + \tau_0(x,1)h(1,\xi) - \int_0^1 \tau_0(x,t) h_t(t,\xi) dt\}, \end{aligned} \tag{33}$$

and differentiation of this leads to the formulas

$$\begin{aligned} \frac{\partial^s H(x,\xi)}{\partial x^s} = -\rho^{-n+s}\Big\{ &\tau_s(x,\xi)[h(\xi-,\xi) - h(\xi+,\xi)] - \tau_s(x,0)h(0,\xi) \\ &+ \tau_s(x,1)h(1,\xi) - \int_0^1 \tau_s(x,t) h_t(t,\xi) dt\Big\} \end{aligned} \tag{34}$$

$$(s = 1, 2, \cdots, n-1).$$

Since $h_t(t, \xi)$ is of bounded variation in t we may apply the second law of the mean to show that the integrals remaining in these formulas are of the form E/ρ. Hence we find that

$$\frac{\partial^s H(x,\xi)}{\partial x^s} = \frac{E}{\rho^{n-s}} \qquad (s = 0, 1, 2, \cdots, n-1). \tag{35}$$

The formula for $\Psi_j(x)$ in (32) may be treated precisely as we have treated that for $H(x, \xi)$. Recalling that $\Phi_j(t)$ is continuous throughout (0, 1) we find from the formulas analogous to (33) and (34) that

$$\frac{d^s\Psi_j(x)}{dx^s} = \rho^{-n+s}\left\{\delta_{0s}\Phi_j(x) + \tau_s(x,0)\Phi_j(0) - \tau_s(x,1)\Phi_j(1) + \frac{E}{\rho}\right\}^* \tag{36}$$

$$(s = 0,1,2,\cdots,n-1).$$

From (35) it is seen directly that the kernel $H(x, \xi)$ of the integral equation (31) possesses a reciprocal kernel $\mathfrak{H}(x, \xi)$ for all values of ρ sufficiently large in absolute value. In "solved" form, then, equation (31) may be written

$$u(x) = \sum_{j=1}^{n} c_j Y_j(x) + \sum_{j=1}^{m} c_j' \Omega_j(x), \tag{37}$$

where

$$\begin{aligned} Y_j(x) &= y_j(x) - \int_0^1 \mathfrak{H}(x,\xi) y_j(\xi)\, d\xi, \\ \Omega_j(x) &= \Psi_j(x) - \int_0^1 \mathfrak{H}(x,\xi)\Psi_j(\xi)\, d\xi. \end{aligned} \tag{38}$$

From the Neumann series for $\mathfrak{H}(x, \xi)$, i.e.

$$\begin{aligned} \mathfrak{H}(x,\xi) = &- H(x,\xi) - \int_0^1 H(x,t)H(t,\xi)\, dt \\ &- \int_0^1 \int_0^1 H(x,t_1)H(t_1,t_2)H(t_2,\xi)\, dt_1 dt_2 - \cdots, \end{aligned}$$

together with formula (35) it is evident that

$$\frac{\partial^s \mathfrak{H}(x,\xi)}{\partial x^s} = \frac{E}{\rho^{n-s}} \qquad (s = 0,1,2,\cdots,n-1). \tag{39}$$

In consequence of this and (36) we see that

$$\frac{d^s\Omega_j(x)}{dx^s} = \rho^{-n+s}\left\{\delta_{0s}\Phi_j(x) + \tau_s(x,0)\Phi_j(0) - \tau_s(x,1)\Phi_j(1) + \frac{E}{\rho}\right\} \tag{40}$$

$$(s = 0,1,2,\cdots.\, n-1),$$

while from (26)

$$\frac{d^s Y_j(x)}{dx^s} = \rho^s e^{\rho\omega_j x}[\omega_j{}^s]. \tag{41}$$

For the evaluation of the characteristic function $u(x)$ from formula (37) it is still necessary to determine the $n+m$ constants c_j and c_j'. For

* $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$.

this determination we have the equations (15b) and (30b) which become on substituting in them the value of $u(x)$ as given by (37)

$$\sum_{k=1}^{n} m_{i,k}\, c_k + \sum_{j=1}^{m} m'_{i,j} c'_j = 0 \qquad (i = 1,2, \cdots, n+m), \tag{42}$$

where

$$m_{i,k} = \begin{cases} L_i(Y_k) - \int_0^1 \alpha_i(\xi) Y_k(\xi)\, d\xi, & \text{when } i \leqq n, \\ -l_p(Y_k), & \text{when } i = n+p; \end{cases}$$

$$m'_{i,j} = \begin{cases} L_i(\Omega_j) - \int_0^1 \alpha_i(\xi)\Omega_j(\xi)\, d\xi, & \text{when } i \leqq n, \\ \delta_{pj} - l_p(\Omega_j), & \text{when } i = n+p. \end{cases} \tag{43}$$

We may assume without loss of generality that the set of operators $L_i(u)$ have been reduced to the normal form [B, p. 382] in which

$$L_i(u) = W_{i0}(u) + W_{i1}(u),$$

where

$$W_{i0}(u) = a_i u^{(k_i)}(0) + \sum_{j=0}^{k_i-1} a_{ij} u^{(j)}(0),$$

$$W_{i1}(u) = b_i u^{(k_i)}(1) + \sum_{j=0}^{k_i-1} b_{ij} u^{(j)}(1),$$

and

$$n - 1 \geqq k_1 \geqq k_2 \geqq \cdots \geqq k_n.$$

On the other hand by the definition of the operators $l_j(u)$ we may assume the subscripts so assigned that

$$l_j(u) = \begin{cases} \text{either } u^{(h_j)}(0), \\ \text{or } \quad u^{(h_j)}(1), \end{cases} \qquad \text{where } n - 1 \geqq h_1 \geqq h_2 \geqq \cdots \geqq h_m.$$

Due to these peculiarities of structure of the operators involved we find readily from (43) with use of (40) that

$$m'_{i,j} = \begin{cases} \rho^{k_i - n} E & \text{for } i \leqq n, \\ [\delta_{pj}] & \text{for } i = n+p. \end{cases}$$

Since $\alpha_i(\xi)$ possesses a derivative which is of bounded variation we find also from (41) that

$$-\int_0^1 \alpha_i(\xi) Y_k(\xi)\, d\xi = \begin{cases} E/\rho & \text{for } k = 1,2, \cdots, \mu, \\ e^{\rho\omega_k} E/\rho & \text{for } k = \mu+1, \cdots, n. \end{cases}$$

Hence it follows that, for $i \leqq n$,

$$m_{i,k} = \begin{cases} (\rho\omega_k)^{k_i}[a_i] & \text{when } k \leqq \mu - 1, \\ (\rho\omega_k)^{k_i}\{[a_i] + e^{\rho\omega_\mu}[b_i]\} & \text{when } k = \mu, \\ (\rho\omega_k)^{k_i}e^{\rho\omega_k}[b_i] & \text{when } k \geqq \mu + 1, \end{cases}$$

while for $i = n+p$,

$$m_{i,k} = \begin{cases} (\rho\omega_k)^{h_p}[a_p'] & \text{when } k \leqq \mu - 1, \\ (\rho\omega_k)^{h_p}\{[a_p'] + e^{\rho\omega_\mu}[b_p']\} & \text{when } k = \mu, \\ (\rho\omega_k)^{h_p}e^{\rho\omega_k}[b_p'] & \text{when } k \geqq \mu + 1, \end{cases}$$

every a_p' and b_p' being either 0 or 1.

If we consider as unknowns in system (42) the quantities c_k'' in place of the quantities c_k where

$$c_k'' = \begin{cases} c_k & \text{when } k = 1, 2, \cdots, \mu, \\ c_k e^{\rho\omega_k} & \text{when } k = \mu + 1, \cdots, n, \end{cases}$$

and remove from the ith equation, $i = 1, 2, \cdots, n$, the factor ρ^{k_i}, the matrix of the resulting system becomes

$$(44) \quad \begin{Bmatrix} [A_{11}] \cdots \{[A_{1\mu}] + e^{\rho\omega_\mu}[B_{1\mu}]\} \cdots [B_{1n}] & E\rho^{-n} \cdots E\rho^{-n} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ [A_{n1}] \cdots \{[A_{n\mu}] + e^{\rho\omega_\mu}[B_{n\mu}]\} \cdots [B_{nn}] & E\rho^{-n} \cdots E\rho^{-n} \\ \rho^{h_1}[A_{11}'] \cdots \rho^{h_1}\{[A_{1\mu}'] + e^{\rho\omega_\mu}[B_{1\mu}']\} \cdots \rho^{h_1}[B_{1n}'] & [\delta_{11}] \cdots [\delta_{1m}] \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \rho^{h_m}[A_{m1}'] \cdots \rho^{h_m}\{[A_{m\mu}'] + e^{\rho\omega_\mu}[B_{m\mu}']\} \cdots \rho^{h_m}[B_{mn}'] & [\delta_{m1}] \cdots [\delta_{mm}] \end{Bmatrix},$$

where the A_{ij}, A_{ij}', B_{ij}, B_{ij}' are constants. The equation $D(\rho) = 0$, where $D(\rho)$ is the determinant of the matrix (44), is precisely the characteristic equation by which the characteristic numbers are determined. Since we suppose ρ a characteristic number it follows that the system (42) admits of a solution in which at least one of the unknowns differs from zero.

The associated matrix for the differential system (15′) is readily shown to be

$$(45) \quad \begin{Bmatrix} [A_{11}] \cdots \{[A_{1\mu}] + e^{\rho\omega_\mu}[B_{1\mu}]\} \cdots [B_{1n}] \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ [A_{n1}] \cdots \{[A_{n\mu}] + e^{\rho\omega_\mu}[B_{n\mu}]\} \cdots [B_{nn}] \end{Bmatrix}.$$

Let the determinant of this matrix be $\Delta(\rho)$. Then $\Delta(\rho) = 0$ is the characteristic equation of the differential system (15′). The condition (vi) that this dif-

ferential system be regular implies that at least one of the cofactors of the elements of the μth column of $\Delta(\rho)$ does not vanish. Due to the structural relationship of the matrices (44) and (45) it is readily seen to follow for $|\rho|$ sufficiently large at least one of the cofactors of the elements of the μth column of $D(\rho)$ is different from zero. Hence the ratios of the unknowns c_k'' and c_j', $k=1, 2, \cdots, n$; $j=1, 2, \cdots, m$, are obtainable from the ratios of the cofactors of a row of the determinant of (44). Substituting in these cofactors the value

$$e^{\rho\omega_\mu} = [e^{2\pi i a/\omega_\mu}],$$

given by (29), we may write

$$(46)\quad c_1 : c_2 : \cdots : c_\mu : e^{\rho\omega_{\mu+1}} c_{\mu+1} : \cdots : e^{\rho\omega_n} c_n : c_1' : \cdots : c_m' = [Q_1] : [Q_2] : \cdots : [Q_n] : \rho^{n-1}[Q_{n+1}] : \cdots : \rho^{n-1}[Q_{n+m}],$$

each Q_j being a constant which may be computed from the elements of (44) and at least one Q_j ($j=1, 2, \cdots, n$) differing from zero. Upon substituting the quantities on the right of (46) for the coefficients in (37) and using (40) and (41) we find the form of the function $u(x)$. The result we state as follows.

THEOREM 4. *If the normalized integral equation satisfies the hypotheses* (iv)-(vi) *the characteristic function $u(x)$ corresponding to the characteristic value $\lambda = \rho^n$ is of the form*

$$u(x) = \sum_{j=1}^{\mu} e^{\rho\omega_j x}[Q_j] + \sum_{j=\mu+1}^{n} e^{\rho\omega_j (x-1)}[Q_j],$$

$$\frac{d^s u(x)}{dx^s} = \rho^s \left\{ \sum_{j=1}^{\mu} e^{\rho\omega_j x}[Q_j \omega_j{}^s] + \sum_{j=\mu+1}^{n} e^{\rho\omega_j (x-1)}[Q_j \omega_j{}^s] \right\} \qquad (s = 1, 2, \cdots, n-1)$$

the quantities Q_j being constants.

10. **The expansion of an arbitrary function.** Let $f(x)$ be any function which is integrable on the interval $(0, 1)$. We consider the following integrals

$$(47)\quad I_N(f) = \frac{1}{2\pi i}\int_{C_N}\int_0^1 \Gamma(x,t,\lambda) f(t)\,dt\,d\lambda,$$

$$I_N'(f) = \frac{1}{2\pi i}\int_{C_N}\int_0^1 \Gamma'(x,t,\lambda) f(t)\,dt\,d\lambda,$$

where Γ and Γ' are the Green's functions defined in §§6 and 7, and C_N is a circle $|\lambda| = R_N$ where $|\lambda_N| < R_N < |\lambda_{N+1}|$.* The radius R_N may be chosen

* If there are infinitely many double characteristic values, N is to take on only those values for which this inequality is possible.

moreover so that C_N for every N sufficiently large has no points in common with any of the small circles c_ν with radii δ and drawn with the points λ_ν as centers. It is known then [T, Theorem 6] that

$$\lim_{N\to\infty}\{I_N(f)-I_N'(f)\}=\begin{cases}0 & \text{uniformly } 0<a\leqq x\leqq b<1,\\ \displaystyle\int_0^1 T_0(t)f(t)dt & \text{for } x=0,\\ \displaystyle\int_0^1 T_1(t)f(t)dt & \text{for } x=1,\end{cases}\tag{48}$$

where the functions $T_0(t)$ and $T_1(t)$ are known and do not depend on $f(x)$.

Now, if we remember the definition of the sets $\{u_\nu, v_\nu\}$, $\{u_\nu', v_\nu'\}$, we may write

$$I_N(f)=\sum\nolimits_N(f),\quad I_M'(f)=\sum\nolimits_M'(f),\tag{49}$$

where

$$\begin{aligned}\sum\nolimits_N(f)&=\sum_{\nu=1}^{N}u_\nu(x)\int_0^1 f(t)v_\nu(t)dt,\\ \sum\nolimits_M'(f)&=\sum_{\nu=1}^{M}u_\nu'(x)\int_0^1 f(t)v_\nu'(t)dt,\end{aligned}\tag{50}$$

M being the number of characteristic values λ_ν' within the circle C_N. By Theorem 3 there exists a positive constant R such that when $R_N>R$ the number of values λ_ν between the circle C_N and the circle $C\colon |\lambda|=R$ is the same as the number of values λ_ν' between these circles. From this it follows that $|N-M|$ is finite, not exceeding the number of characteristic values λ_ν within the circle C.

Consider now the identity

$$\sum\nolimits_M'(f)-\sum\nolimits_N'(f)=\pm\sum_{\nu=N}^{M}\frac{1}{2\pi i}\int_{c_\nu'}\int_0^1\Gamma'(x,t,\lambda)f(t)dtd\lambda.\tag{51}$$

Since

$$|\Gamma'(x,t,\lambda)|<A|\lambda|^{-(n-1)/n},$$

where A is a constant [T, §18], it follows that each of the integrals on the right of (51) converges uniformly to zero as $N\to\infty$. We readily draw from this the result which we formulate as follows.

THEOREM 5. *If the normalized integral equation satisfies the hypotheses* (iv)-(vi) *then the series*

$$\sum\nolimits_\infty(f)\quad \textit{and}\quad \sum\nolimits_\infty'(f)$$

are equiconvergent for every integrable function $f(x)$ in the interior of the interval (0, 1), *i.e.*

$$\lim_{N\to\infty} \{\sum\nolimits_N(f) - \sum\nolimits'_N(f)\} = 0 \text{ uniformly for } 0 < a \leqq x \leqq b < 1.$$

At the end points $x=0$ and $x=1$ the difference of the series converges to

$$\int_0^1 T_0(t)f(t)dt \quad \text{and} \quad \int_0^1 T_1(t)f(t)dt$$

*respectively, where $T_0(t)$ and $T_1(t)$ are known and do not depend on $f(x)$.**

The conclusions reached above enable us now to state also the following facts [T, §35]:

THEOREM 6. *If the normalized integral equation satisfies hypotheses* (iv)-(vi) *with $n \geqq 2$, then when $|\lambda-\lambda_\nu| \geqq \delta$ and $|\lambda|$ is sufficiently large its kernel and its resolvent kernel admit of the absolutely and uniformly convergent expansions*

$$\Re(x,t,\lambda) = \sum_{\nu=1}^{\infty} \Re_\nu(x,t,\lambda), \quad K(x,t) = -\sum_{\nu=1}^{\infty} \Re_\nu(x,t,0),$$

where $\Re_\nu(x, t, \lambda)$ is the principal part of $\Re(x, t, \lambda)$ at the pole $\lambda=\lambda_\nu$. The solution of the non-homogeneous equation is then given by

$$u(x) = f(x) - \lambda \sum_{\nu=1}^{\infty} \int_0^1 \Re_\nu(x,t,\lambda)f(t)dt,$$

the series converging uniformly and absolutely for every integrable function $f(x)$.

If all the characteristic values are simple, these expansions are, more explicitly,

$$\Re(x,t,\lambda) = \sum_{\nu=1}^{\infty} \frac{u_\nu(x)v_\nu(t)}{\lambda-\lambda_\nu}, \qquad K(x,t) = \sum_{\nu=1}^{\infty} \frac{u_\nu(x)v_\nu(t)}{\lambda_\nu},$$

$$u(x) = (fx) - \lambda \sum_{\nu=1}^{\infty} \frac{u_\nu(x)}{\lambda-\lambda_\nu} \int_0^1 f(t)v_\nu(t)dt.$$

An analogous theorem has been stated also in the case where $n=1$ [T, §35].

---

* Let $\{u_\nu', v_\nu'\}$ be the set of fundamental functions of the problem $L(u)+\lambda u=0$, $L_i(u)=\int_0^1 \alpha_i(\xi)u(\xi)d\xi$. On setting $\sum_N'(f)=\sum_{\nu=1}^N u_\nu'(x)\int_0^1 f(t)v_\nu'(t)dt$, it is readily proved that the series $\sum_N(f)$ and $\sum_N'(f)$ are equiconvergent on the whole of (0, 1), the end points inclusive.

BROWN UNIVERSITY,
PROVIDENCE, R. I.;
UNIVERSITY OF WISCONSIN,
MADISON, WIS.

ON APPROXIMATION TO AN ARBITRARY FUNCTION OF A COMPLEX VARIABLE BY POLYNOMIALS*

BY
J. L. WALSH

1. **Introduction.** In the theory of functions of a complex variable there are many problems relating to uniform approximation (or what is essentially the same thing, to uniform expansion) of an arbitrary function by means of polynomials, which have not yet been completely solved. Let us mention a few examples.

When can a function $f(z)$ defined on a closed point set M be uniformly approximated on M by polynomials in z? Of what point sets M is it true that every function continuous on M can be approximated in this way? Of what closed point sets M is it true that every function $f(z)$ continuous on M and analytic at every interior point of M can be uniformly approximated? What conformal maps can be approximately performed by means of polynomials, with an error uniformly and arbitrarily small?

The present paper does not give a complete answer to any of these questions, but gives new results connected with each of them, particularly when the point sets M in question consist of Jordan curves or arcs which do not necessarily refrain from intersecting. Hitherto only such configurations as consist of non-intersecting curves have been treated, with reference to uniform approximation of functions. The results of the present paper are the most general results which can be obtained in answer to the questions raised, concerning point sets M which are composed of a finite number of Jordan arcs, curves, and regions.

If a point set M extends to infinity and is closed, the only function $f(z)$ continuous on M which can be uniformly approximated on M by polynomials is a constant. Hence throughout this paper we consider, unless the contrary is explicitly stated, only point sets which do not extend to infinity.

The results which are to be obtained here are proved by the use of modern work in conformal mapping—a theory which has yielded most of the recent results on approximation of functions of a complex variable.

2. **Statement of some results obtained; connection with the literature.** Let us indicate more specifically the advance made in the present paper. Runge's theorem is classical:

---

* Presented to the Society, December 28, 1927; received by the editors in November, 1927.

If the function $f(z)$ is analytic in a closed Jordan region, then in that closed region $f(z)$ can be uniformly expanded in a series of polynomials in z.*

The following results have more recently been established:†

THEOREM I. *If the function $f(z)$ is analytic interior to a Jordan curve and continuous in the corresponding closed region, then in that closed region $f(z)$ can be uniformly expanded in a series of polynomials in z.*

THEOREM II. *If the function $f(z)$ is continuous on a Jordan arc, then on that arc $f(z)$ can be uniformly expanded in a series of polynomials in z.*

The following more general theorems, among others, are to be proved here:

THEOREM III. *If the two Jordan curves C_1 and C_2 are exterior to each other except for a single common point $z=\alpha$, and if the functions $f_1(z)$ and $f_2(z)$ are respectively analytic interior to C_1 and C_2 and continuous in the corresponding closed regions, and if $f_1(\alpha)=f_2(\alpha)$, then there exists a series of polynomials in z which converges uniformly to the sum $f_1(z)$ in the closed interior of C_1 and uniformly to the sum $f_2(z)$ in the closed interior of C_2.*

THEOREM IV. *If the bounded point set M consists of a finite number of Jordan arcs and does not divide the plane, then an arbitrary function $f(z)$ continuous on M is developable on M in a uniformly convergent series of polynomials in z.*

3. **A theorem on the conformal mapping of variable regions.** For later use in the study of expansions we require the following theorem:

THEOREM V. *In the z-plane let the Jordan curves C_1 and C_2 lie exterior to each other, except that they have the single point $A(z=\alpha)$ in common. Let B_1 and B_2 be fixed points of C_1 and C_2 respectively. Let the Jordan curve K_n, $n=1, 2, \cdots$, contain the interiors of C_1 and C_2 and consist (at least for n sufficiently great) of an arc of C_1 which contains B_1, an arc of C_2 which contains B_2, and of two arcs of the circle with center A and radius $1/n$. If $z'=\omega_n(z)$ denotes the function that maps the interior of K_n onto the interior of C_1 so*

---

* We emphasize the fact that the region here is not the most general simply connected region of the plane; proof that the theorem breaks down in this more general case will be found on pp. 432–433 of the paper referred to in the next following footnote. Misunderstanding of this fact caused the statement of an erroneous theorem in the Encyklopädie, II C4, Bieberbach, *Neuere Untersuchungen über Funktionen von komplexen Variablen*, p. 49.

† By application, suggested by Carathéodory, of a theorem on conformal mapping due to Courant. See Walsh, Mathematische Annalen, vol. 96 (1926), pp. 430–436; 437–450.

that a fixed point D interior to C_1 and a fixed direction at that point remain invariant, then we have

$$\lim_{n\to\infty} \omega_n(z) = \omega(z) = \begin{cases} z, & z \text{ on or interior to } C_1, \\ \alpha, & z \text{ on or interior to } C_2, \end{cases} \tag{1}$$

uniformly on the closed point set consisting of the curves C_1 and C_2 and their interiors.

We do not prove (1) directly, but find it more convenient to map the interiors of C_1 and of K_n onto the interior of the unit circle Γ in the τ-plane, in such a way that D is transformed into the origin and the given direction at D into the same direction at $\tau=0$. Denote by $\tau=\phi(z)$, $\phi_n(z)$ the respective mapping functions. Then (1) is entirely equivalent to the uniform approach to a limit:

$$\lim_{n\to\infty} \phi_n(z) = \begin{cases} \phi(z), & z \text{ on or interior to } C_1, \\ \phi(\alpha), & z \text{ on or interior to } C_2. \end{cases} \tag{2}$$

For the two functions $z'=\omega_n(z)$ and $\tau=\phi_n(z)$ are obtained from each other by the transformation $\tau=\phi(z')$ and its inverse, both of which are continuous transformations in the closed region consisting of C_1 and its interior. We proceed, then, to the proof of (2).

The result

$$\lim_{n\to\infty} \phi_n(z) = \phi(z), \; z \text{ on or interior to } C_1',$$

uniformly, when C_1' is a fixed Jordan curve interior to C_1, follows from a theorem due to Bieberbach.* But by results of Lindelöf, points interior

* Göttinger Nachrichten, 1913, pp. 552–560. The names of Carathéodory (Mathematische Annalen, vol. 73 (1912), pp. 107–144) and of Courant (Hurwitz-Courant, *Funktionentheorie*, Berlin, 1925, pp. 385–387) should also be mentioned here.

We apply the theorem of Bieberbach, loc. cit., p. 557, although he considers the mapping of regions K_n on *variable* regions, and with the auxiliary conditions (in our notation) $\phi_n(0)=0$, $\phi_n'(0)=1$, where D is the point $z=0$. This involves no real difficulty, for we need merely modify our $\phi_n(z)$ and $\phi(z)$ by proper multiplicative constants to bring about this change; the new functions map the interior of K_n on variable circles whose Kern (Carathéodory) is the transform of the interior of C_1. Uniform convergence of these modified functions $\phi_n(z)$ results from Bieberbach's theorem, and implies the uniform convergence of our original functions $\phi_n(z)$ as used in the present paper.

The connection of this entire subject, the conformal mapping of variable regions, with Lebesgue's work on harmonic functions (Rendiconti del Circolo Matematico di Palermo, vol. 24 (1907), pp. 371–402) seems not to have been pointed out in the literature, but we could in fact use his results instead of those of Bieberbach. We outline briefly the method. By the results of Lebesgue (pp. 398–399), the Green's function (as defined by Osgood, *Funktionentheorie*, Leipzig, 1912, p. 630) for the interior of K_n approaches uniformly in the closed interior of C_1 (strictly speaking, the single point D is an exception) the Green's function for the interior of C_1. The sequence of functions conjugate to the Green's functions for the interior of K_n approaches likewise the function conjugate to the Green's function for the interior of C_1 and uniformly in any closed region interior to C_1, when this conjugate function is defined properly for our auxiliary conditions $\phi_n'(0)>0$, $\phi'(0)>0$. It follows that the mapping function $\phi_n(z)$ converges uniformly to the limit $\phi(z)$ in any closed region interior to C_1.

to C_1 and K_n and within a distance δ of the boundary are transformed by $\tau=\phi(z)$, $\phi_n(z)$ into points within a distance $\eta(\delta)$ of the boundary of Γ, where $\eta(\delta)$ is independent of n and approaches zero with δ.* Moreover, a point P interior to C_1 or K_n which can be connected to a point Q of the boundary of C_1 or K_n by a Jordan arc (interior to C_1 or K_n) of diameter less than δ is transformed into points P', P_n interior to Γ whose distances from the points Q', Q_n corresponding to Q are respectively less than $\theta(\delta)$, where $\theta(\delta)$ is independent of n and approaches zero with δ.

Let an arbitrary positive ϵ be given. Choose δ so that $\theta(\delta)<\epsilon/6$. Choose the curve C_1' so that every point R of C_1' is within a distance $\delta/2$ of C_1 and is interior to C_1, and so that every point S on or within C_1 but not within C_1' can be joined interior to C_1 by a Jordan arc of diameter less than $\delta/2$ to some point of C_1'. Choose also $N>4/\delta$. Then an arbitrary point S on or within C_1 but not within C_1', even such a point S in the neighborhood of A, can be joined interior to C_1 by a Jordan arc of diameter less than δ to some point Q of C_1 and of K_n for $n>N$.

Denote by Q', R', S', and Q_n, R_n, S_n the transforms of Q, R, S respectively under the transformations $\tau=\phi(z)$, $\tau=\phi_n(z)$. Then if the points R and S previously considered can be joined within C_1 by a Jordan arc of diameter less than $\delta/2$, the points R and S can be joined within C_1 by a Jordan arc of diameter less than δ to some point Q (the same point Q for both R and S) of C_1 and of K_n, so we have

$$R'Q'<\frac{\epsilon}{6},\quad Q'S'<\frac{\epsilon}{6},\quad R_nQ_n<\frac{\epsilon}{6},\quad Q_nS_n<\frac{\epsilon}{6},$$

and hence we have

$$R'S'<\frac{\epsilon}{3},\quad R_nS_n<\frac{\epsilon}{3},$$

uniformly for all R, S, and $n>N$.

But by choosing n sufficiently large we have by Bieberbach's result, $R'R_n<\epsilon/3$ uniformly for all R on C_1' and for all sufficiently large n. If S is an arbitrary point on or within C_1 but not within C_1', there *exists* some point R of C_1' which can be joined within C_1 to S by a Jordan arc of diameter less than $\delta/2$. Thus we have, by combining the inequalities already derived,

$$S'S_n=\left|\phi(z)-\phi_n(z)\right|<\epsilon$$

uniformly for all sufficiently large n and for all points S on or within C_1 but not within C_1'. This, together with the result of Bieberbach,

* Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46 (1915), No. 4, *Sur un principe général de l'analyse*, pp. 1–35; especially pp. 15–18.

$$\lim_{n\to\infty} \phi_n(z) = \phi(z), \text{ uniformly for all } z \text{ on or within } C_1',$$

yields immediately the first part of (2),

$$\lim_{n\to\infty} \phi_n(z) = \phi(z), \text{ uniformly for all } z \text{ on or within } C_1.$$

If each point of the curve C_1' is within a distance $\delta/2$ of the curve C_1, the transform J_1' of C_1' is everywhere within a distance $\eta(\delta/2+1/n)$ of the circle Γ, and the transform J_2' of C_2 is a Jordan curve lying outside of J_1', under the transformation $\tau=\phi_n(z)$. Every point of J_2' not a point of Γ lies exterior to J_1', and within some circle of radius $\theta(2/n)$, for *all* points of C_2 not points of K_n can be connected by a Jordan curve interior to K_n whose diameter is less than $2/n$ to a *single* point Q of K_n. We have already proved, however,

$$\lim_{n\to\infty} \phi_n(\alpha) = \phi(\alpha),$$

from which follows, since $\phi_n(\alpha)$ lies on J_2', the second part of (2), and (2) is completely established.

We remark that Theorem V can be extended to include the case of several (even an infinity of suitably restricted) Jordan curves C_2, each having but a single point in common with C_1; the proof already given holds without essential change. In the theorem as stated it is not essential that the curves K_n should be composed partly of arcs of circles. It is sufficient (although even this is not necessary for the proof) that the arcs of K_n not part of C_1 or C_2 lie in a circle whose center is A and whose radius approaches zero as n becomes infinite.

The writer has had the privilege of discussing the proof of Theorem V with Dr. Karl Löwner of Berlin, and hereby expresses his indebtedness to him for aid in formulating that proof.

4. **Approximation on two point sets having a single common point.** As a matter of general notation, if we denote a Jordan curve by C_i, the closed point set which consists of the curve and its interior will be denoted by $\overline{C}_i$.

Our most important tool in later work is a consequence of Theorem V:

THEOREM VI. *Let the Jordan curves C_1 and C_2 lie exterior to each other except for a single point A, $z=\alpha$, which they have in common. Let the sequence of polynomials $\{P_n'(z)\}$ converge to a function $F_1(z)$ defined throughout $\overline{C}_1$ or more generally on any point set which contains A and is contained in $\overline{C}_1$. Similarly let the sequence of polynomials $\{P_n''(z)\}$ converge to a function $F_2(z)$ defined throughout $\overline{C}_2$ or more generally on any point set which contains A and is contained in $\overline{C}_2$. Then if $F_1(\alpha)=F_2(\alpha)$, there exists a sequence of polynomials $\{P_n(z)\}$ such that*

$$\lim_{n\to\infty} [P_n(z) - P_n'(z)] = 0, \text{ uniformly for } z \text{ in } \overline{C}_1,$$

(3)

$$\lim_{n\to\infty} [P_n(z) - P_n''(z)] = 0, \text{ uniformly for } z \text{ in } \overline{C}_2.$$

It follows that in $\overline{C}_1$ the points of convergence of the sequences $\{P_n(z)\}$ and $\{P_n'(z)\}$ are the same, likewise point sets of uniform convergence; in case the former sequence converges with a certain sum, the latter sequence converges with the same sum, and conversely. A similar remark applies to the sequences $\{P_n(z)\}$ and $\{P_n''(z)\}$ in $\overline{C}_2$.

There can easily be proved by the classical method of Runge* a result similar to Theorem VI for the case that $\overline{C}_1$ and $\overline{C}_2$ have no point in common. Thus, if in this case the sequences of polynomials $\{P_n'(z)\}$ and $\{P_n''(z)\}$ converge in $\overline{C}_1$ and $\overline{C}_2$ respectively or on point sets contained in those regions, there exists a sequence of polynomials $\{P_n(z)\}$ satisfying the requirements (3). This result can also be proved easily by direct application of Theorem VI.† We shall apply this result later, in the proof of Theorem VII.

Theorem V is to be applied in the proof of Theorem VI. The function $\omega_n(z)$ is analytic interior to K_n, and continuous in the region $\overline{K}_n$, which includes of course the point set $\overline{C}_1+\overline{C}_2$. Hence by Theorem I the function $\omega_n(z)$ is in $\overline{K}_n$ uniformly developable in series of polynomials, and therefore the limit of this function, $\omega(z)$, approached uniformly, is in $\overline{C}_1+\overline{C}_2$ likewise uniformly developable in a series of polynomials.

Let us suppose for the sake of simplicity that $F_1(\alpha)=F_2(\alpha)=0$; this assumption involves no loss of generality. Choose an arbitrary sequence $\{\epsilon_n\}$ of positive numbers whose limit is zero. Since $\omega(z)$ is uniformly developable on $\overline{C}_1+\overline{C}_2$ in a series of polynomials, there exists a polynomial $Q_n'(z)$ such that

$$|Q_n'(z) - P_n'(z)| < \frac{\epsilon_n}{2}, \quad z \text{ on } \overline{C}_1,$$

$$|Q_n'(z) - P_n'(\alpha)| < \frac{\epsilon_n}{2}, \quad z \text{ on } \overline{C}_2.$$

---

* See for example Montel, *Séries des polynomes à une variable complexe*, Paris, 1910, Ch. IV.

† We sketch this suggested proof. Choose a point A, $z=\alpha$, not in $\overline{C}_1$ or $\overline{C}_2$; apply Theorem VI to the polynomial sequences $\{(z-\alpha)P_n'(z)\}$ and $\{(z-\alpha)P_n''(z)\}$ considered in new Jordan regions $\overline{C}_1'$ and $\overline{C}_2'$, which have the single common point A and which contain $\overline{C}_1$ and $\overline{C}_2$ respectively. We thus arrive at a sequence of polynomials which we write $\{P_n''(z)\}$ and consider in $\overline{C}_1$ and $\overline{C}_2$. But in $\overline{C}_1$ and $\overline{C}_2$ the function $1/(z-\alpha)$ can be expanded uniformly in a series of polynomials, and hence the sequence $\{P_n''(z)/(z-\alpha)\}$ may be replaced by a sequence of polynomials $\{P_n(z)\}$ having in $\overline{C}_1$ and $\overline{C}_2$ the properties required.

The polynomial $Q_n'(z)$ may be taken equal to $P_n'[\Omega(z)]$, where $\Omega(z)$ is a polynomial which approximates $\omega(z)$ sufficiently closely on $\overline{C}_1+\overline{C}_2$.

Similarly there exists a polynomial $Q_n''(z)$ such that

$$|Q_n''(z) - P_n''(z)| < \frac{\epsilon_n}{2}, \quad z \text{ on } \overline{C}_2,$$

$$|Q_n''(z) - P_n''(\alpha)| < \frac{\epsilon_n}{2}, \quad z \text{ on } \overline{C}_1.$$

Then the polynomial $Q_n(z)=Q_n'(z)+Q_n''(z)$ satisfies the inequalities

$$|Q_n(z) - P_n''(\alpha) - P_n'(z)| < \epsilon_n, \ z \text{ on } \overline{C}_1,$$
$$|Q_n(z) - P_n'(\alpha) - P_n''(z)| < \epsilon_n, \ z \text{ on } \overline{C}_2,$$

so that by virtue of

$$\lim_{n\to\infty} P_n'(\alpha) = F_1(\alpha) = 0, \quad \lim_{n\to\infty} P_n''(\alpha) = F_2(\alpha) = 0,$$

the sequence $\{Q_n(z)\}$ satisfies the requirements of Theorem VI.

5. **Approximation on m given point sets.** Theorem III follows immediately from Theorems I and VI. A generalization of Theorem VI which considers not merely two but an arbitrary number of Jordan curves will turn out to be convenient:

THEOREM VII. *Let the point set M be composed of the closed interiors of the Jordan curves $C_1, C_2, \cdots, C_m$; we suppose that these curves lie exterior to each other, except that C_i and C_j $(i, j=1, 2, \cdots, m, i\neq j)$ may have a single point $z=\alpha_{ij}$ (but no more) in common. We suppose also that the system of curves C_i divides the plane into precisely $m+1$ distinct regions. Then if the sequence of polynomials $\{P_n^{(i)}(z)\}$, $i=1, 2, \cdots, m$, converges to a limit function $F_i(z)$ defined on the entire point set $\overline{C}_i$ or a part of it, and if finally $F_i(\alpha_{ij})=F_j(\alpha_{ij})$ for all points α_{ij} which exist,* then there exists a sequence of polynomials $\{P_n(z)\}$ such that*

$$\lim_{n\to\infty} [P_n(z) - P_n^{(i)}(z)] = 0, \text{ uniformly on } \overline{C}_i, i = 1,2, \cdots, m. \tag{4}$$

From (4) follows, as in Theorem VI, that in $\overline{C}_i$ the points of convergence and limits of the sequences $\{P_n(z)\}$ and $\{P_n^{(i)}(z)\}$ are the same, likewise point sets of uniform convergence, and so on.

* Here there are two interpretations possible, (a) that when α_{ij} exists, both functions $F_i(z)$ and $F_j(z)$ are required to be defined and equal there, or (b) that when α_{ij} exists and when $F_i(\alpha_{ij})$ and $F_j(\alpha_{ij})$ are both defined, then these two functional values shall be equal. We understand here the *former* interpretation.

Theorem VII has already been established for the case $m=2$. Let us proceed by induction, assuming the theorem true for $m=\mu-1$ and proving the theorem for $m=\mu$. At least one of the curves C_i, let us say for definiteness C_1, has not more than *one* point in common with the totality of other curves. Otherwise every curve C_i would possess at least two such points, there would exist a new (i.e., distinct from the C_i) Jordan curve consisting only of arcs of the curves C_i, and the μ curves C_i would separate the plane into more than $\mu+1$ regions. Hence there exists a Jordan curve K which is exterior to C_1, except that K may have a single point in common with C_1, and such that $\overline{C}_i$ belongs to $\overline{K}$ for $i=2, 3, \cdots, m=\mu$. Theorem VII now follows, on the one hand by application of Theorem VII for $m=\mu-1$ to the point set $\overline{C}_2+\overline{C}_3+\cdots+\overline{C}_\mu$, and on the other hand by application to C_1 and K either of Theorem VI or of the result (method of Runge) similar to Theorem VI, according as C_1 and K have or have not a common point.

The following theorem, of which Theorem III is a special case, results from Theorem VII without further argument:

THEOREM VIII. *If the point set M satisfies the conditions of Theorem* VII *and if the function** *$f(z)$ is analytic interior to C_i and continuous on M, then on M the function $f(z)$ can be developed in a uniformly convergent series of polynomials.*

6. **A general result on approximation.** Our most general theorem will now be proved; both Theorems VIII and IV are included here as special cases.

THEOREM IX. *Let the point set M consist of the closed interiors of the Jordan curves $C_1, C_2, \cdots, C_m$, and of the Jordan arcs $B_1, B_2, \cdots, B_n$. Suppose the curves $C_i (i\neq j)$ and the arcs $B_1, B_2, \cdots, B_n$ are all exterior to the curve C_j except that possibly a single point (but no more) of C_i or B_k may coincide with a point of C_j. Suppose finally that the curves $C_1, \cdots, C_m, B_1, \cdots, B_n$ do not separate the plane into more than $m+1$ distinct regions. Then an arbitrary function analytic in the interior points of M and continuous on M can be developed uniformly on M in series of polynomials.*

We suppose the notation such—and this involves no loss of generality—that no point of an arc B_i other than an end point is a point of C_k or of $B_j (i\neq j)$.

There can be drawn Jordan curves $C_{m+1}, C_{m+2}, \cdots, C_{m+n}$ so that B_i belongs to $\overline{C}_{m+i}$, so that no two of the curves C_i $(i=1, 2, \cdots, m+n)$

*Of course, not necessarily a monogenic analytic function.

have more than a single point in common, so that these curves lie exterior to each other except for these common points, and so that the curves C_1 $\cdots$, C_{m+n} do not divide the plane into more than $m+n+1$ distinct regions. This construction is possible by enclosing the arc B_i in a Jordan curve C_{m+i}, drawn sufficiently close to B_i and having on it no point of M other than the end points of B_i, and having no point in common with other arcs B_j or with a curve $C_k (k=1, 2, \cdots, m+n, k \neq m+i)$ except such points as B_i has in common with those curves. Theorem IX then follows immediately, by application of Theorems I, II, and VII.

The kind of function $f(z)$ considered in Theorem IX is the most general function which can be uniformly expanded on M in a series of polynomials. For the sum of such a uniformly convergent series is continuous on M and analytic in the interior points of M.

Theorem IX breaks down if we omit from the hypothesis the requirement that the curves $C_1, \cdots, C_m, B_1, \cdots, B_n$ shall not separate the plane into more than $m+1$ distinct regions. Proof of this fact is essentially to be found in the discussion to be given later, in §8.

7. **Remarks on conformal mapping.** Theorem IX has interesting applications to conformal mapping. If we have in the z-plane the point set M of that theorem and in the w-plane a second point set M' consisting of Jordan curves $C_1', C_2', \cdots, C_m'$, and of Jordan arcs $B_1', B_2', \cdots, B_n'$, having the other properties mentioned in the theorem, and finally if the point sets M and M' can be mapped on each other one-to-one and continuously without exception and conformally so far as concerns the interiors of C_i and C_i', then this mapping can be accomplished with any desired degree of approximation by means of polynomials, from the z-plane to the w-plane and the reverse. For the mapping functions satisfy all the requirements of Theorem IX.

Two arbitrary point sets M and M' having the properties mentioned in the theorem and which are topologically equivalent cannot always be mapped on each other in this way. For example, a circle C_1 with four Jordan arcs B_i abutting on it cannot be so mapped on an arbitrary circle C_1' with four Jordan arcs abutting on it. For the four points common to C_1 and B_i cannot in general be transformed into the four points common to C_1' and B_i'; the most general conformal map of a circle onto another circle involves *three* real parameters. But in every case if M and M' satisfy the requirements of Theorem IX, if they are topologically equivalent, and if C_i has no more than three points in common with $C_1+C_2+\cdots+C_{i-1}+C_{i+1}+\cdots+C_m+B_1+\cdots+B_n$, then M and M' can be mapped on each other continuously and one-to-one, conformally so far as concerns the in-

teriors of C_i and C_i', and this transformation can be accomplished with an arbitrary degree of approximation by means of a polynomial in z or w.

To the point set M of Theorem IX there may be adjoined without altering the truth of the theorem an arbitrary number of additional points P, provided these new points have at most a finite number of limit points and that the entire point set, M plus the new points, is closed. The property still holds, that if $f(z)$ is analytic in the interior points of M and continuous on the entire point set, then on that entire point set $f(z)$ can be uniformly expanded in a series of polynomials. The details here are easy and are omitted.* This remark has obvious applications to conformal mapping by means of polynomials.

8. **Approximation by rational functions.** We return to the questions raised in §1. The following theorem is extremely easy to prove, but perhaps so near to these questions that it is worth stating:

Let M be a limited, closed point set. Then a necessary and sufficient condition that every function $f(z)$ analytic on M be uniformly developable on M in a series of polynomials, is that M be the complementary set (with respect to the entire plane) of an infinite region.

Here the function $f(z)$ is not required to be a monogenic analytic function, but is merely required to be (regular-) analytic at every point of M.

This condition on M is sufficient, for there can be drawn—for instance, by the help of a network of squares—a finite number of Jordan curves all exterior to each other such that each curve has on and within it only points of analyticity of $f(z)$, and such that every point of M is interior to one of these curves. The fact that $f(z)$ is developable follows by Runge's results.

The condition on M is necessary. For let us assume the contrary, that there exists a point P not a point of M but which cannot be joined to the point at infinity by a curve which does not meet M. Then all points Q which can be joined to P by curves which do not meet M form a region R' all of whose boundary points are points of M. A region R exists which is simply-connected, which contains all points of R', and whose boundary points are all points of M. Then the function $f(z)=1/(z-p)$, where P is the point

---

* Compare a remark due to Szegö and the present writer, Walsh, loc. cit., p. 443, footnote. Many other points sets N, in particular certain of them containing an infinity of Jordan arcs B, can be made essentially to come under this theorem. For addition of new points to N may enable us to construct thereby a point set M satisfying the requirements of Theorem IX. A function continuous on N is easily extended in definition so as to be continuous also on M. Of course we add in the general case, where an arbitrary function $f(z)$ is concerned, no interior points to N, nor do we increase the number of Jordan curves of N.

$z=p$, is analytic in all points of M but cannot be developed on M uniformly in a series of polynomials.*

The subject of approximation of arbitrary functions by means of *rational functions* instead of by polynomials is an interesting one which has received little attention in the literature. All the questions raised in the first part of the present paper can properly be raised for approximation by means of rational functions instead of by means of polynomials, and these questions are still, in their ultimate generality, unanswered. One extremely simple theorem is perhaps worth mentioning, although the proof presents no difficulty. This theorem and the next one stated hold even for points sets which extend to infinity, provided that to the plane a single point at infinity be adjoined:

Let M be an arbitrary closed point set of the plane. If the function $f(z)$ is analytic in every point of M, then $f(z)$ can be expanded on M in a uniformly convergent series of rational functions of z.

Deeper lying results can be established by the use of conformal mapping, and the methods of the present paper can be used here to some extent. As an illustration of results that can be obtained in this way, we state without proof at this time the following theorem:

If the point set M consists of two Jordan curves C_1 and C_2 which have only a single point in common, then an arbitrary function $f(z)$ continuous on M can be uniformly expanded on M in a series of rational functions. If the fixed points P_1, P_2, P_3 are chosen arbitrarily, one in each of the three regions into which M divides the plane, then each rational function of the series can be chosen so that it has no poles except in P_1, P_2, and P_3.

---

* See Walsh, loc. cit., p. 434.

If M is a limited, closed point set, then a necessary and sufficient condition that a given function $f(z)$ analytic on M be uniformly developable on M in a series of polynomials, is that $f(z)$ or its analytic extension (i.e. all analytic extensions which can be obtained from points of M along paths which do not elsewhere meet M) be analytic in every point which cannot be connected with the point at infinity by a curve which does not meet M.

This condition can be otherwise expressed. All points which can be connected with the point at infinity by curves which do not meet M form a region B. The condition is then *that $f(z)$ or its analytic extension be analytic in every point of the point set complementary (with respect to the entire plane) to the region B.*

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

TRANSFORMATIONS OF NETS*

BY
V. G. GROVE

INTRODUCTION

It is the purpose of this paper to extend some of the ideas relating to transformations of surfaces and conjugate nets to transformations of an arbitrary net. When two nets and the congruence of lines joining corresponding points are so related that the developables of the congruence cut the sustaining surfaces of the nets in the curves of the net, the nets and the congruences will be said to be in *relation C*. Either net will be said to be a C transform of the other. If the given nets are conjugate nets, then the transformation C becomes a transformation F.† If corresponding points of two nets in relation C separate harmonically the focal points on the line joining corresponding points, we shall call the given nets K_{-1} *transforms*. If the given nets are conjugate nets, then the transformation K_{-1} becomes the transformation K of Koenig.‡ We shall consider these and related transformations in some detail.

1. THE DIFFERENTIAL EQUATIONS

Let

$$y^{(k)} = y^{(k)}(u,v), \quad z^{(k)} = z^{(k)}(u,v) \qquad (k = 1,2,3,4)$$

be the parametric equations of surfaces S_y and S_z but assume that neither surface is a focal surface of the congruence G of lines g joining the points y and z. Suppose that the parametric nets N_y and N_z on S_y and S_z and the congruence G are in relation C. Evidently the tangents to the curves $u = \text{const.}$ on S_y and S_z are coplanar and similarly the tangents to $v = \text{const.}$ are coplanar. The functions $y^{(k)}$ and $z^{(k)}$ will satisfy a system of differential equations of the form§

---

* Presented to the Society, December 31, 1926; received by the editors in October, 1927.

† L. P. Eisenhart, *Transformations of Surfaces*, Princeton University Press, 1923, p. 34. Hereafter referred to as Eisenhart, *Surfaces*.

‡ Eisenhart, *Surfaces*, p. 57.

§ G. M. Green, *Memoir on the general theory of surfaces and rectilinear congruences*, these Transactions, vol. 20 (1919), p. 149. Hereafter referred to as Green, *Surfaces*.

$$
\begin{aligned}
y_{uu} &= a^{(11)}y_u + b^{(11)}y_v + c^{(11)}y + d^{(11)}z,\\
y_{uv} &= a^{(12)}y_u + b^{(12)}y_v + c^{(12)}y + d^{(12)}z,\\
y_{vv} &= a^{(22)}y_u + b^{(22)}y_v + c^{(22)}y + d^{(22)}z,\\
z_u &= my_u + py + qz,\\
z_v &= n'y_v + p'y + q'z.
\end{aligned}
\tag{1}
$$

The coefficients of system (1) satisfy the following integrability conditions*:

$$
\begin{aligned}
&a_u^{(12)} - a_v^{(11)} + a^{(12)}b^{(12)} - a^{(22)}b^{(11)} + c^{(12)} = -md^{(12)},\\
&b_u^{(12)} - b_v^{(11)} + a^{(12)}b^{(11)} + (b^{(12)} - a^{(11)})b^{(12)} - b^{(11)}b^{(22)} - c^{(11)} = n'd^{(11)},\\
&c_u^{(12)} - c_v^{(11)} + a^{(12)}c^{(11)} + (b^{(12)} - a^{(11)})c^{(12)} - b^{(11)}c^{(22)} = p'd^{(11)} - pd^{(12)},\\
&d_u^{(12)} - d_v^{(11)} + a^{(12)}d^{(11)} + (b^{(12)} - a^{(11)})d^{(12)} - b^{(11)}d^{(22)} = q'd^{(11)} - qd^{(12)};
\end{aligned}
\tag{2a}
$$

$$
\begin{aligned}
&a_u^{(22)} - a_v^{(12)} + a^{(22)}a^{(11)} + (b^{(22)} - a^{(12)})a^{(12)} - b^{(12)}a^{(22)} + c^{(22)} = -md^{(22)},\\
&b_u^{(22)} - b_v^{(12)} + a^{(22)}b^{(11)} - a^{(12)}b^{(12)} - c^{(12)} = n'd^{(12)},\\
&c_u^{(22)} - c_v^{(12)} + a^{(22)}c^{(11)} + (b^{(22)} - a^{(12)})c^{(12)} - b^{(12)}c^{(22)} = p'd^{(12)} - pd^{(22)},\\
&d_u^{(22)} - d_v^{(12)} + a^{(22)}d^{(11)} + (b^{(22)} - a^{(12)})d^{(12)} - b^{(12)}d^{(22)} = q'd^{(12)} - qd^{(22)};
\end{aligned}
\tag{2b}
$$

$$
\begin{aligned}
&-m_v + (n' - m)a^{(12)} + p' = -q'm,\\
&n_u' + (n' - m)b^{(12)} - p = qn',\\
&p_u' - p_v + (n' - m)c^{(12)} = qp' - pq',\\
&q_u' - q_v + (n' - m)d^{(12)} = 0.
\end{aligned}
\tag{2c}
$$

If the fourth and fifth equation of system (1) are solved for y_u and y_v, and if the resulting two equations are differentiated with respect to u and v, we obtain the following equivalent system:

* Green, *Surfaces*, p. 150.

$$
\begin{aligned}
z_{uu} &= \alpha^{(11)}z_u + \beta^{(11)}z_v + \gamma^{(11)}z + \delta^{(11)}y,\\
z_{uv} &= \alpha^{(12)}z_u + \beta^{(12)}z_v + \gamma^{(12)}z + \delta^{(12)}y,\\
z_{vv} &= \alpha^{(22)}z_u + \beta^{(22)}z_v + \gamma^{(22)}z + \delta^{(22)}y,\\
y_u &= \mu z_u + \pi z + \kappa y,\\
y_v &= \nu' z_v + \pi' z + \kappa' y,
\end{aligned} \tag{3}
$$

wherein

$$
\alpha^{(11)} = a^{(11)} + \frac{p}{m} + \frac{m_u}{m} + q, \qquad \beta^{(11)} = \frac{m}{n'}b^{(11)},
$$

$$
\delta^{(11)} = -m\left[a^{(11)}\frac{p}{m} + b^{(11)}\frac{p'}{n'} + \left(\frac{p}{m}\right)^2 - c^{(11)} - \frac{\partial}{\partial u}\left(\frac{p}{m}\right)\right], \tag{3a}
$$

$$
\alpha^{(12)} = a^{(12)} + \frac{m_v}{m}, \quad \beta^{(12)} = b^{(12)} + \frac{n_u'}{n'},
$$

$$
\begin{aligned}
\delta^{(12)} &= -m\left[a^{(12)}\frac{p}{m} + b^{(12)}\frac{p'}{n'} - \frac{pp'}{mn'} - c^{(12)} - \frac{\partial}{\partial v}\left(\frac{p}{m}\right)\right]\\
&= -n'\left[b^{(12)}\frac{p'}{n'} + a^{(12)}\frac{p}{m} - \frac{pp'}{mn'} - c^{(12)} - \frac{\partial}{\partial u}\left(\frac{p'}{n'}\right)\right],
\end{aligned} \tag{3b}
$$

$$
\alpha^{(22)} = \frac{n'}{m}a^{(22)}, \beta^{(22)} = b^{(22)} + \frac{p'}{n'} + \frac{n_v'}{n'} + q',
$$

$$
\delta^{(22)} = -n'\left[a^{(22)}\frac{p}{m} + b^{(22)}\frac{p'}{n'} + \left(\frac{p'}{n'}\right)^2 - c^{(22)} - \frac{\partial}{\partial v}\left(\frac{p'}{n'}\right)\right], \tag{3c}
$$

$$
\mu = \frac{1}{m}, \quad \nu' = \frac{1}{n'}, \quad \kappa = -\frac{p}{m}, \quad \kappa' = -\frac{p'}{n'}, \quad \pi = -\frac{q}{m}, \quad \pi' = -\frac{q'}{n'}. \tag{3d}
$$

The differential equation of the asymptotic curves* on S_y is

$$
d^{(11)}du^2 + 2d^{(12)}dudv + d^{(22)}dv^2 = 0, \tag{4}
$$

and the asymptotic curves on S_z are given by the equation

$$
\delta^{(11)}du^2 + 2\delta^{(12)}dudv + \delta^{(22)}dv^2 = 0. \tag{5}
$$

The surface S_y is therefore a developable if $d^{(11)}d^{(22)} - (d^{(12)})^2 = 0$. A similar statement holds for S_z. We shall assume hereafter that neither S_y nor S_z is developable. If $d^{(12)} = 0$ the parametric net on S_y is conjugate; if $d^{(11)} = d^{(22)} = 0$ the parametric net is asymptotic. Similar statements hold for the parametric net on S_z.

* Green, *Surfaces*, p. 151.

2. The focal nets of G

Any point P on g is defined by an expression of the form

$$P = y + \theta z.$$

As $y(z)$ describes a curve $v=v(u)$ on $S_y(S_z)$, g generates a ruled surface. If this surface is a developable, the point dP/du lies on g. Imposing this condition we find that the developables are given by the differential equation

$$(n' - m)dudv = 0 \tag{6}$$

and the focal points by

$$mn'\theta^2 + (m + n')\theta + 1 = 0. \tag{7}$$

The focal points of g are therefore

$$\begin{aligned} \tau &= my - z, \\ \tau' &= n'y - z. \end{aligned} \tag{8}$$

We shall call the nets N_τ and $N_{\tau'}$ *the focal nets of* G. If the focal surfaces S_τ and $S_{\tau'}$ are not developable, the nets N_τ and $N_{\tau'}$ are conjugate. Equation (6) shows that the developables are indeterminate if $m-n'=0$. In this case we shall call N_y and N_z *radial transforms*. Unless explicitly stated to the contrary, we shall hereafter assume that N_y and N_z are not radial transforms.

3. The transformations L and L'

From the last two equations of system (1) we find that the points of intersection of corresponding tangents to the curves of nets in relation C are defined by the expressions

$$\begin{aligned} r &= y_u + \frac{p}{m}y = \mu(z_u - qz), \\ s &= y_v + \frac{p'}{n'}y = \nu'(z_v - q'z). \end{aligned} \tag{9}$$

The line h of intersection of the tangent planes to S_y and S_z at corresponding points y and z generates a congruence H. We may in the usual way show that the differential equation of the net corresponding to the developables of the congruence H is

$$\begin{aligned} (md^{(11)}\delta^{(12)} - n'd^{(12)}\delta^{(11)})du^2 + [m(d^{(11)}\delta^{(22)} + d^{(12)}\delta^{(12)}) \\ - n'(d^{(22)}\delta^{(11)} + d^{(12)}\delta^{(12)})]dudv - (n'd^{(22)}\delta^{(12)} - md^{(12)}\delta^{(22)})dv^2 = 0. \end{aligned} \tag{10}$$

The focal points of h are defined by the expressions

$$P_{1,2} = r + \theta_{1,2}s,$$

where $\theta_{1,2}$ are the roots of

$$(11)\quad m(d^{(22)}\delta^{(12)} - d^{(12)}\delta^{(22)})\theta^2 + [n'(d^{(22)}\delta^{(11)} - d^{(12)}\delta^{(12)}) - m(d^{(11)}\delta^{(22)} - d^{(12)}\delta^{(12)})]\theta - n'(d^{(11)}\delta^{(12)} - d^{(12)}\delta^{(11)}) = 0.$$

The harmonic invariant of (4) and (10) is

$$(m - n')(d^{(11)}d^{(22)} - d^{(12)2})\delta^{(12)} = 0.$$

Hence *the congruence H will be harmonic to S_y if and only if G is conjugate to S_z. Similarly H will be harmonic to S_z if and only if G is conjugate to S_y.*

We shall call nets N_y and N_z *L transforms* or in *relation L* if they are in relation C and if they correspond to the developables of H. Nets N_y and N_z in relation C will be L transforms if

$$(12)\quad md^{(11)}\delta^{(12)} - n'd^{(12)}\delta^{(11)} = n'd^{(22)}\delta^{(12)} - md^{(12)}\delta^{(22)} = 0.$$

Hence *if N_y and N_z are in relation F they are L transforms. If N_y and N_z are asymptotic nets, they are L transforms.* Suppose N_y and N_z are neither conjugate nor asymptotic. Then from (12) we obtain

$$\frac{d^{(11)}d^{(22)}}{d^{(12)2}} = \frac{\delta^{(11)}\delta^{(22)}}{\delta^{(12)2}}.$$

Hence *the tangents to the curves of non-asymptotic nets in relation L form the same cross ratios with the asymptotic tangents associated with the surfaces on which the nets lie.* In particular *if the tangents to the curves of a net on a surface form a constant cross ratio with the asymptotic tangents, any L transform of the net has the same property, the constant ratio being the same for all nets in relation L*.*

From (1) we see that the focal points of h lie on the tangents to the curves of nets in relation C if and only if

$$(13)\quad d^{(22)}\delta^{(12)} - d^{(12)}\delta^{(22)} = d^{(11)}\delta^{(12)} - d^{(12)}\delta^{(11)} = 0.$$

We shall call such nets N_y and N_z *L' transforms* or nets in *relation L'*. Hence *if N_y and N_z are F transforms they are L' transforms; if they are asymptotic nets in relation C they are L' transforms; if they are in relation C and the asymptotic curves on the surfaces on which the nets lie correspond then the nets are L' transforms.*

---

* This statement of the theorem includes the case in which the nets are conjugate.

From equations (13) we find that

$$\frac{d^{(11)}d^{(22)}}{d^{(12)^2}} = \frac{\delta^{(11)}\delta^{(22)}}{\delta^{(12)^2}}.$$

Hence *the tangents to the curves of non-asymptotic nets in relation L' form the same cross ratio with the tangents to the asymptotic curves on the surfaces on which the nets lie.* In particular *if the tangents to the curves of a net on a surface form a constant cross ratio with the asymptotic tangents, any L' transform has the same property, the constant ratio being the same for all nets in relation L'.**

From (12) and (13) we see that *nets in relation C are both L and L' transforms if they are in relation F or are both asymptotic nets.* From (11) we find that *the focal points of h corresponding to non-conjugate nets N_v and N_z are indeterminate if and only if N_v and N_z are radial transforms and the asymptotic curves on S_v and S_z correspond.* In this case *the net* (10) *is indeterminate. If N_v and N_z are conjugate nets the focal points of h are indeterminate if these nets are radial transforms and if the asymptotic nets on S_v and S_z correspond.*

4. Coaxial nets in relation C

Let the curves of the two nets N_v and N_z be union curves of the congruence G. We shall call such nets *coaxial* nets since they have the same axis. From the first and the third equations of (1) and (3) it follows that the nets N_v and N_z in relation C are coaxial nets if and only if $b^{(11)} = a^{(22)} = \beta^{(11)} = \alpha^{(22)} = 0$. From (3a) and (3c) it follows that *if a net N_v is in relation C to its axis congruence G, any C transform N_z of N_v by the same congruence G has G for its axis congruence. N_v and N_z are coaxial nets in relation C.*

We may readily verify that the curve $v = \text{const.}$ on S_v is a plane curve if and only if the invariant

$$(14) \qquad G_v = b^{(11)}(a^{(11)}d^{(11)} + b^{(11)}d^{(12)} + qd^{(11)} + d_u^{(11)}) - d^{(11)}(a^{(11)}b^{(11)} + b^{(11)}b^{(12)} + b_u^{(11)})$$

vanishes; the curve $u = \text{const.}$ on S_v is plane if and only if

$$(15) \qquad G_v' = a^{(22)}(b^{(22)}d^{(22)} + a^{(22)}d^{(12)} + q'd^{(22)} + d_v^{(22)}) - d^{(22)}(a^{(22)}b^{(22)} + a^{(22)}a^{(12)} + a_v^{(12)})$$

vanishes. Similar expressions G_z and G_z' may be written for the curves $v = \text{const.}$ and $u = \text{const.}$ on S_z. From (14) and (15) it follows that *if N_v*

* This statement of the theorem includes the case in which the nets are conjugate.

*and N_z are coaxial nets in relation C they are nets of plane curves.** Moreover it is readily proved that *if N_y is a net of plane curves and N_y and N_z are coaxial they are in relation C, and N_z is also a net of plane curves.*

Suppose that N_y and N_z are in relation C to the axis congruence of N_y and that N_z has for one (or both) of its component families one (or both) of the families of asymptotic curves of S_z. Then that family of curves consists of a set of rulings on the ruled surface S_z.

The ray of the point y with respect to the net N_y joins the points

$$\begin{aligned} \rho_y &= y_u - \frac{1}{d^{(11)}}(b^{(12)}d^{(11)} - b^{(11)}d^{(12)})y, \\ \sigma_y &= y_v - \frac{1}{d^{(22)}}(a^{(12)}d^{(22)} - a^{(22)}d^{(12)})y. \end{aligned} \tag{16}$$

The ray of the point z with respect to N_z joins the points

$$\begin{aligned} \rho_z &= z_u - \frac{1}{\delta^{(11)}}(\beta^{(12)}\delta^{(11)} - \beta^{(11)}\delta^{(12)})z, \\ \sigma_z &= z_v - \frac{1}{\delta^{(22)}}(\alpha^{(12)}\delta^{(22)} - \alpha^{(22)}\delta^{(12)})z. \end{aligned} \tag{17}$$

Suppose that ρ_y and ρ_z are distinct. The line joining ρ_y and ρ_z intersects g in the point

$$\begin{aligned} P_\rho = &\left[p + \frac{m}{d^{(11)}}(b^{(12)}d^{(11)} - b^{(11)}d^{(12)}) \right] y \\ &+ \left[q - \frac{1}{\delta^{(11)}}(\beta^{(12)}\delta^{(11)} - \beta^{(11)}\delta^{(12)}) \right] z. \end{aligned} \tag{18}$$

Similarly $\sigma_y\sigma_z$ intersects g in the point P_σ defined by

$$\begin{aligned} P_\sigma = &\left[p' + \frac{n'}{d^{(22)}}(a^{(12)}d^{(22)} - a^{(22)}d^{(12)}) \right] y \\ &+ \left[q' - \frac{1}{\delta^{(22)}}(\alpha^{(12)}\delta^{(22)} - \alpha^{(22)}\delta^{(12)}) \right] z. \end{aligned} \tag{19}$$

The point P_ρ will coincide with the focal point τ' of g if and only if

$$n'\left[q - \frac{1}{\delta^{(11)}}(\beta^{(12)}\delta^{(11)} - \beta^{(11)}\delta^{(12)}) \right] = - p - \frac{m}{d^{(11)}}(b^{(12)}d^{(11)} - b^{(11)}d^{(12)}).$$

* G. M. Green, *Nets of space curves*, these Transactions, vol. 21 (1920), p. 231. Hereafter referred to as Green, *Nets*. See also G. M. Green, *Projective differential geometry of one-parameter families of curves*, etc., American Journal of Mathematics, vol. 38 (1916), p. 304.

Using the expressions for $\beta^{(11)}$ and $\beta^{(12)}$ from (3a) and (3b) this condition may be reduced to

$$b^{(11)}\frac{d^{(12)}}{d^{(11)}} = b^{(11)}\frac{\delta^{(12)}}{\delta^{(11)}}.$$

Similarly the line $\sigma_y\sigma_z$ meets g in the focal point τ if and only if

$$a^{(22)}\frac{d^{(12)}}{d^{(22)}} = a^{(22)}\frac{\delta^{(12)}}{\delta^{(22)}}.$$

Let us call the points ρ_y, ρ_z, τ' *focal points of the first rank*, and σ_y, σ_z, τ *focal points of the second rank. It follows therefore that the lines joining corresponding focal points of the tangents to the curves of nets in relation C with respect to a congruence intersect the lines of the congruence in the focal points of the same rank if*

(a) *the nets are coaxial, or if*

(b) *the nets are in relation F,* or if*

(c) *the asymptotic curves on S_y and S_z correspond.*

5. The derived congruences of G. The relation R

We shall call the tangent to $C_u(C_v)$ on the focal surface $S_{\tau'}(S_\tau)$ the minus first (first) derived congruence of G. We find that the minus first derived line of g joins the point τ' to the point

$$T_y^{-1} = \frac{\tau_u' - q\tau'}{n' - m} = y_u - b^{(12)}y. \tag{20}$$

Similarly the first derived line of g joins τ to the point

$$T_y^{+1} = \frac{\tau_v - q'\tau}{m - n'} = y_v - a^{(12)}y. \tag{21}$$

Comparing (20) and (16), we find that the minus first derived line of g intersects the tangent to C_u on S_y in the focal point if and only if $b^{(11)}d^{(12)} = 0$.

Similarly the first derived line of g meets the tangent to C_v on S_y in the focal point if and only if $a^{(22)}d^{(12)} = 0$.

Hence *the minus first (first) derived line of g intersects the tangent to $C_u(C_v)$ on S_y in the focal point on this tangent if g lies in the osculating plane to $C_u(C_v)$ at y, or if N_y is a conjugate net.* A similar statement may be made of N_z. Hence *the derived lines of g meet the tangents to the curves of nets in relation C*

* Eisenhart, *Surfaces*, p. 94, exercise 20.

in the focal points of these tangents if (a) *the nets are in relation F or if* (b) *they are coaxial nets.*

We may readily verify that the asymptotic nets on $S_{r'}$ and S_r are defined by the differential equations

$$(22)\qquad \begin{aligned}(p' - n'_v + n'q')du^2 - (n' - m)b^{(11)}dv^2 = 0,\\ (p - m_u + mq)dv^2 - (m - n')a^{(22)}du^2 = 0,\end{aligned}$$

respectively. Therefore, if the line g in relation C to N_y and N_z lies in the osculating plane to $C_u(C_v)$ on S_y or S_z, the focal surface $S_{r'}(S_r)$ is developable. *If N_y and N_z are coaxial nets in relation C the focal surfaces of G are both developables, the generators forming the lines of G.*

The points T_y^{-1} and T_y^{+1} defined by (20) and (21) have an interesting interpretation. Any point r on the tangent to C_u at y is defined by an expression of the form

$$r = y_u - \lambda y.$$

As y moves along C_v the tangent to C_u on S_y generates a ruled surface $R^{(u)}$ and r traces out a curve on $R^{(u)}$. A point on the tangent to this curve is

$$\eta = r_v - ar = (a^{(12)} - a)y_u - (b^{(12)} - \lambda)y_v + (c^{(12)} - \lambda_v + \alpha\lambda)y + d^{(12)}z.$$

Hence the tangent to the curve traced by r as y moves along C_v lies in the plane determined by g, r and y if and only if

$$\lambda = b^{(12)}.$$

This value of λ determines the point

$$T_y^{-1} = y_u - b^{(12)}y.$$

The point T_y^{+1} has a similar interpretation.

The line $T_y^{-1}T_y^{+1}$ is called *the line in relation R to g* with respect to N_y*. The points

$$(23)\qquad T_z^{-1} = z_u - \beta^{(12)}z, \quad T_z^{+1} = z_v - \alpha^{(12)}z$$

may be similarly defined. If N_y is a conjugate net the points T_y^{-1} and T_y^{+1} are the Laplace transforms of y. We may state our results as follows:

The derived lines of g meet the tangents to the curves of non-conjugate nets in relation C to G in points which determine the line g' in relation R to g. The derived lines of g meet the tangents to the curves of conjugate nets in relation F to G in the Laplace transforms of the surface point y.

---

* Green, *Surfaces*, p. 86, and footnote p. 87.

The tangent to C_v at T_v^{-1} on the surface generated by this point cuts g in the point

$$(a^{(12)}b^{(12)} + c^{(12)} - b_v^{(12)})y + d^{(12)}z. \tag{24}$$

Similarly the tangent to C_u at T_v^{+1} cuts g in the point

$$(a^{(12)}b^{(12)} + c^{(12)} - a_u^{(12)})y + d^{(12)}z. \tag{25}$$

The points determined by (24) and (25) coincide if $d^{(12)}=0$ or if $a_u^{(12)}-b_v^{(12)}=0$*. In case $d^{(12)}\neq 0$ and $a_u^{(12)}-b_v^{(12)}=0$ the developables of the congruence of lines g' in relation R to g correspond to a conjugate net.

Comparing the expressions (20) and (21) with (9), we find that the line h of intersection of corresponding tangent planes to S_y and S_z is in relation R to g with respect to N_y if and only if

$$b^{(12)} + \frac{p}{m} = 0, \quad a^{(12)} + \frac{p'}{n'} = 0. \tag{26}$$

Similarly h will be in relation R to g with respect to N_z if and only if

$$\beta^{(12)} - q = 0, \quad \alpha^{(12)} - q' = 0. \tag{27}$$

From (3b) and (2c) we find that

$$\alpha^{(12)} - q' = \frac{n'}{m}\left(a^{(12)} + \frac{p'}{n'}\right),$$

$$\beta^{(12)} - q = \frac{m}{n'}\left(b^{(12)} + \frac{p}{m}\right).$$

Hence *if N_y, N_z, and G are in relation C and if H is in relation R to G with respect to N_y it is in relation R to G with respect to N_z.*

From (16) and (17) we note that the focal points ρ_y and ρ_z coincide if and only if

$$m(b^{(12)}d^{(11)} - b^{(11)}d^{(12)}) + pd^{(11)} = 0,$$

$$\beta^{(12)}\delta^{(11)} - \beta^{(11)}\delta^{(12)} - q\delta^{(11)} = 0.$$

These conditions may be reduced by means of (3a) and (3b) to

$$b^{(12)} + \frac{p}{m} = b^{(11)}\frac{d^{(12)}}{d^{(11)}}, \quad b^{(12)} + \frac{p}{m} = b^{(11)}\frac{\delta^{(12)}}{\delta^{(11)}}. \tag{28}$$

---

* V. G. Grove, *A theory of a general net on a surface*, these Transactions, vol. 28 (1926), p. 495, footnote.

Similarly the focal points σ_y and σ_z coincide if and only if

$$a^{(12)} + \frac{p'}{n'} = a^{(22)}\frac{d^{(12)}}{d^{(22)}}, \quad a^{(12)} + \frac{p'}{n'} = a^{(22)}\frac{\delta^{(12)}}{\delta^{(22)}}. \tag{29}$$

We shall call nets N_y and N_z *coradial nets* if they have the same rays. Hence *coaxial nets in relation C are coradial if and only if h and g are in relation R. If the nets are in relation F they are coradial if and only if the two derived lines of g intersect h.* In order for nets which are non-conjugate and not coaxial to be coradial it is necessary that the asymptotic curves on the sustaining surfaces correspond.

We may readily verify that the lines in relation R to g with respect to N_y and N_z intersect h in the points

$$\left(a^{(12)} + \frac{p'}{n'}\right)\boldsymbol{r} - \left(b^{(12)} + \frac{p}{m}\right)\boldsymbol{s}, \quad n'\left(a^{(12)} + \frac{p'}{n'}\right)\boldsymbol{r} - m\left(b^{(12)} + \frac{p}{m}\right)\boldsymbol{s} \tag{30}$$

respectively. These points coincide if N_y and N_z are radial transforms. Suppose that neither of the points (30) coincide with $\boldsymbol{r}$ or $\boldsymbol{s}$. The cross ratio of the points $\boldsymbol{r}$, $\boldsymbol{s}$ and the points (30) is m/n'. Hence *if the lines in relation R to g meet h in distinct points, the cross ratio of these points and the points* $\boldsymbol{r}$ *and* $\boldsymbol{s}$*, and the cross ratio of the points y, z, τ, τ′ are equal.*

Similarly we may prove that if N_y and N_z are L' transforms the rays of N_y and N_z cut h in points forming with $\boldsymbol{r}$ and $\boldsymbol{s}$ the same cross ratio that y and z form with the focal points of g.

6. The transformations K_R

We shall call nets N_y and N_z in relation C K_R *transforms* if

$$\frac{\partial^2}{\partial u \partial v}\log R = 0, \tag{31}$$

wherein the cross ratio $(y, \tau', z, \tau) = m/n' = R$. If N_y and N_z are in relation F and are K_{-1} transforms, they are K transforms in the sense of Koenig* since corresponding surface points separate the focal points of g harmonically.

From (3b) we find that

$$\alpha_u^{(12)} - \beta_v^{(12)} = a_u^{(12)} - b_v^{(12)} + \frac{\partial^2}{\partial u \partial v}\log R.$$

Hence *if the line in relation R to g with respect to a net is harmonic to* S_y *the same property is enjoyed by any* K_R *transform of the given net.*

* Comptes Rendus, vol. 113 (1891), p. 1022. Eisenhart, *Surfaces*, p. 58.

Suppose that N_y and N_z are K_{-1} transforms. We find from the integrability conditions (2c) and $m+n'=0$, that

$$\frac{\partial}{\partial v}\log m + 2a^{(12)} + \frac{p'}{n'} = q',$$
$$\frac{\partial}{\partial v}\log n' + 2b^{(12)} + \frac{p}{m} = q. \tag{32}$$

Differentiating (32) with respect to u and v respectively and subtracting, we find that

$$2(a_u^{(12)} - b_v^{(12)}) + \frac{\partial}{\partial u}\left(\frac{p'}{n'}\right) - \frac{\partial}{\partial v}\left(\frac{p}{m}\right) + q_v - q_u' = 0.$$

But from (2c)

$$q_u' - q_v = (n' - m)d^{(12)};$$

and from (3b)

$$\frac{\partial}{\partial u}\left(\frac{p'}{n'}\right) - \frac{\partial}{\partial v}\left(\frac{p}{m}\right) = (\nu' - \mu)\delta^{(12)}.$$

We have, therefore, the theorem of Koenig*: *K transforms have equal point invariants.*

Let now N_y and N_z be coradial nets in relation F. From (28) and (29) and (2c) it follows that

$$\frac{\partial^2}{\partial u \partial v}\log R = 0.$$

Hence *if N_y and N_z are coradial nets in relation F, they have equal point invariants.*†

The ray curves of N_y form a conjugate net‡ if

$$\frac{\partial}{\partial u}\left(a^{(12)} - \frac{a^{(22)}d^{(12)}}{d^{(22)}}\right) - \frac{\partial}{\partial v}\left(b^{(12)} - \frac{b^{(11)}d^{(12)}}{d^{(11)}}\right) = 0. \tag{33}$$

The ray curves of N_z form a conjugate net if

$$\frac{\partial}{\partial u}\left(\alpha^{(12)} - \frac{\alpha^{(22)}\delta^{(12)}}{\delta^{(22)}}\right) - \frac{\partial}{\partial v}\left(\beta^{(12)} - \frac{\beta^{(11)}\delta^{(12)}}{\delta^{(11)}}\right) = 0. \tag{34}$$

Let N_y and N_z be L transforms. Equation (34) may be written

---

* Comptes Rendus, vol. 113 (1891), p. 1022.
† Eisenhart, *Surfaces*, p. 68, exercise 16.
‡ See V. G. Grove, loc. cit.

$$\frac{\partial}{\partial u}\left(a^{(12)} - a^{(22)}\frac{d^{(12)}}{d^{(22)}}\right) - \frac{\partial}{\partial v}\left(b^{(12)} - b^{(11)}\frac{d^{(12)}}{d^{(11)}}\right) = -\frac{\partial^2}{\partial u \partial v}\log R.$$

Hence *if two nets N_y and N_z are L transforms in the relation of a transformation K_R, the ray curves of N_z will form a conjugate system if and only if the ray curves of N_y form a conjugate system.*

7. Pencils of conics in the tangent planes

Consider the points

$$r' = y_u - \lambda y, \quad s' = y_v - \mu y, \tag{35}$$

lying on the tangent lines at y to the curves of the net N_y. If use be made of (9), (35) may be written

$$r' = \mathbf{r} - \left(\lambda + \frac{p}{m}\right)y, \quad s' = \mathbf{s} - \left(\mu + \frac{p}{m}\right)y.$$

The pencil of conics tangent to the parametric tangents at r' and s' is therefore

$$2k_1x_2x_3 = \left[x_1 + \left(\lambda + \frac{p}{m}\right)x_2 + \left(\mu + \frac{p'}{n'}\right)x_3\right]^2, \quad x_4 = 0, \tag{36}$$

the tetrahedron of reference being y, $\mathbf{r}$, $\mathbf{s}$, z.

Similarly the pencil of conics tangent at

$$r'' = z_u - \lambda' z = m\mathbf{r} - (\lambda' - q)z,$$
$$s'' = z_v - \mu' z = n'\mathbf{s} - (\mu' - q')z,$$

referred to the same tetrahedron is

$$2k_2x_2x_3 = \left[\frac{\lambda' - q}{m}x_2 + \frac{\mu' - q'}{n'}x_3 + x_4\right]^2, \quad x_1 = 0. \tag{37}$$

The conics (36) and (37) determine two involutions on the line h. These involutions will be identical if and only if

$$(p + \lambda m)^2(\mu' - q')^2 = (p' + \mu n')^2(\lambda' - q)^2. \tag{38}$$

Suppose that the points r', s', r'', s'' are the focal points on the lines on which they lie. From (16) and (17) it follows that

$$\lambda = b^{(12)} - \frac{b^{(11)}d^{(12)}}{d^{(11)}}, \quad \lambda' = \beta^{(12)} - \frac{\beta^{(11)}\delta^{(12)}}{\delta^{(11)}},$$
$$\mu = a^{(12)} - \frac{a^{(22)}d^{(12)}}{d^{(22)}}, \quad \mu' = \alpha^{(12)} - \frac{\alpha^{(22)}\delta^{(12)}}{\delta^{(22)}}. \tag{39}$$

Suppose now that N_y and N_z are L' transforms. Equations (38) under conditions (39) may be written

$$\left(p + mb^{(12)} - \frac{mb^{(11)}d^{(12)}}{d^{(11)}}\right)^2 \left(\frac{p' + n'a^{(12)}}{m} - \frac{n'a^{(22)}d^{(12)}}{md^{(22)}}\right)^2$$
$$= \left(p' + n'a^{(12)} - \frac{n'a^{(22)}d^{(12)}}{d^{(22)}}\right)^2 \left(\frac{p + mb^{(12)}}{n'} - \frac{mb^{(11)}d^{(12)}}{d^{(11)}}\right)^2.$$

Hence

$$m^2 - n'^2 = 0.$$

We may state our results as follows: *Let the parametric nets N_y and N_z be in relation L'. The pencils of conics tangent to the parametric tangents at their focal points determine the same involution on h if and only if L' is K_{-1}.* If the parametric nets are in relation F, we have the theorem of Eisenhart*:

The two pencils of conics tangent to the parametric tangents to the curves of nets in relation F determine the same involution on the line of intersection of corresponding tangent planes if and only if the nets are in relation K.

Let N_y and N_z be K_{-1} transforms in relation L'. The two conics (36) and (37) under conditions (39) have a point in common if and only if

$$k_1 = m^2 k_2 + 2\left(\lambda + \frac{p}{m}\right)\left(\mu + \frac{p'}{n'}\right). \tag{40}$$

The pencil of quadrics determined by these conics under condition (40) is

$$\begin{aligned} &x_1^2 + \left(\lambda + \frac{p}{m}\right)^2 x_2^2 + \left(\mu + \frac{p'}{n'}\right)^2 x_3^2 + m^2 x_4^2 + 2\left(\lambda + \frac{p}{m}\right) x_1 x_2 \\ &+ 2\left(\mu + \frac{p'}{n'}\right) x_1 x_3 + 2k_3 x_1 x_4 - 2\left[\left(\frac{p}{m} + \lambda\right)\left(\frac{p'}{m'} + \mu\right)\right. \\ &\left. + m^2 k_2\right] x_2 x_3 - 2m\left(\lambda + \frac{p}{m}\right) x_2 x_4 + 2m\left(\mu + \frac{p'}{n'}\right) x_3 x_4 = 0, \end{aligned} \tag{41}$$

the tetrahedron of reference being y, r, s, z. The two quadrics of the pencil (41) tangent to the line g are determined by the parameter values $k_3 = \pm m$. These values of k_3 are the parameter values of the cones of the pencil provided all the quadrics of the pencil are not cones. The coördinates of the vertices of the cones are for

---

* L. P. Eisenhart, *Conjugate systems with equal point invariants*, Annals of Mathematics, (2), vol. 18 (1916), p. 14.

$$(42)\qquad \begin{aligned} k_3 &= m, \quad \left[m^2 k_2,\ 0,\ 2m\left(\lambda + \frac{p}{m}\right),\ k_1\right], \\ k_3 &= -m, \quad \left[m^2 k_2,\ 2m\left(\mu + \frac{p'}{n'}\right),\ 0,\ k_1\right]. \end{aligned}$$

We readily verify that the first vertex of (42) lies on $\sigma_y\tau$, and the second vertex on $\rho_y\tau'$. The pencil (41) cuts the line g in an involution whose double points are the focal points of g. We may state our results as follows:

If N_y and N_z are K_{-1} transforms in relation L', any two of the conics (36) *and* (37) *with values of λ and μ given by* (39) *meeting on h determine a pencil of quadrics which cut g in an involution whose double points are the focal points of g. The two cones of the pencil are the quadrics tangent to g at the focal points of g. The vertices of the cones are on the lines joining corresponding focal points on the parametric tangents.*

If in the above theorem the restriction that N_y and N_z be L' transforms be removed, and the condition of being coaxial be imposed on the nets the remainder of the theorem is true. Also *let N_y and N_z be coaxial nets in relation C. It follows that the pencil of conics tangent to the parametric tangents at their focal points determine the same involution if and only if C is K_{-1}.*

Let now N_y and N_z be non-conjugate nets in relation C not necessarily L'. The wording of the above theorem for parametric values $\lambda = b^{(12)}$, $\mu = a^{(12)}$ then holds.

The vertices of the cones in question lie in this case on the derived lines of g.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

ALLGEMEINE EIGENSCHAFTEN DER CANTORSCHEN KOHÄRENZEN*

VON
MIRON ZARYCKI

Die Ableitung einer Menge A ist die Menge der Häufungspunkte von A. Ich bezeichne dieselbe mit A^d. G. Cantor† nannte die Menge der in A enthaltenen Häufungspunkte von A die (erste) Kohärenz von A. Ich bezeichne sie mit A^k.

In der gegenwärtigen Abhandlung beweise ich einige allgemeine Eigenschaften der Kohärenz von beliebigen Mengen. Alle Sätze erscheinen als Folgen von drei unabhängigen "Axiomen" und ihre Beweise werden mit Hilfe der Grundoperationen der Algebra der Logik durchgeführt. Um die Richtigkeit der als Axiome angenommenen Formeln zu beweisen, musste ich zuerst einige Eigenschaften des Begriffs der Ableitung besprechen.

1. Einige Eigenschaften des Begriffs der Ableitung einer Menge

1. Ich bezeichne den "Raum," in dem alle zu besprechenden Mengen enthalten sind, mit C und die Komplementärmenge von A mit A^c. Dann gelten für beliebige in C enthaltene Mengen folgende Formeln:

I_d: $$(A + B)^d = A^d + B^d;$$

II_d: $$C^d = C;$$

III_d: $$A^{dd} \subset A^d.$$

2. Auf Grund dieser bekannten Formeln beweise ich nun folgende Sätze:

1_d: $$\textit{Aus } A \subset B \textit{ folgt } A^d \subset B^d;$$

2_d: $$A^d - B^d \subset (A - B)^d;$$

3_d: $$A^{cdc} \subset A^{dcdc} \subset A^d;$$

4_d: $$A^{cdc} \subset A^{cdcd} \subset A^d.$$

Der Satz 1_d folgt unmittelbar aus der Formel I_d. Um den Satz 2_d zu beweisen, benutzen wir die evidente Formel $A \subset AB^c + B$. Daraus folgt nach 1_d und I_d:

$$A^d \subset (AB^c + B)^d = (AB^c)^d + B^d$$

---

* Presented to the Society, February 25, 1928; received by the editors in March, 1927.
† G. Cantor, Acta Mathematica, 1885.

und nach beiderseitigen Multiplikation mit B^{dc}:

$$A^d B^{dc} \subset (AB^c)^d B^{dc} \subset (AB^c)^d.^*$$

Setzen wir nun in der Formel 2_d: $A = C$. Man erhält dann nach II_d und 2_d:

$$B^{dc} = C^d D^{dc} \subset (CB^c)^d = B^{cd}$$

und daraus durch die Kontraposition: $B^{cdc} \subset B^d$, oder wenn man jetzt die sonst allgemeine Menge B mit A bezeichnet: $A^{cdc} \subset A^d$. Nach 1_d und III_d folgt daraus:

$$(\alpha) \qquad A^{cdcd} \subset A^{dd} \subset A^d.$$

Die letzte Formel gilt für jede Menge, also auch für A^c, und man bekommt daher $A^{ccdcd} \subset A^{cd}$, oder $A^{dcd} \subset A^{cd}$, und daraus:

$$(\beta) \qquad A^{cdc} \subset A^{dcdc}.$$

In der Formel $A^{cdc} \subset A^d$ setzen wir jetzt A^{cd} anstatt A und es kommt: $A^{cdcdc} \subset A^{cdd} \subset A^{cd}$, und daraus:

$$(\gamma) \qquad A^{cdc} \subset A^{cdcd}.$$

Indem man in (γ) A^c anstatt A setzt, bekommt man $A^{dc} \subset A^{dcd}$, und daraus:

$$(\delta) \qquad A^{dcdc} \subset A^d.$$

Die Relationen 3_d und 4_d sind somit bewiesen.

2. Das Axiomensystem

1. Die Grundlage nachstehender Untersuchungen bildet das folgende Axiomensystem:

I_k: $\quad A^k + B^k \subset (A + B)^k$;

II_k: $\quad A^k \subset A$;

III_k: $\quad A^{ckck} = A^{kckc}$.

2. Die Relation I_k folgt aus der Formel I_d. Man bekommt nämlich

$$A^k + B^k = AA^d + BB^d,$$

und

$$(A+B)^k = (A+B)(A+B)^d = (A+B)(A^d+B^d) = AA^d + BB^d + AB^d + BA^d;$$

also:

$$A^k + B^k \subset (A + B)^k.$$

---

* Es sei bemerkt, dass die Menge AB^c mit der Menge $A - B$ identisch ist.

Die Relation II_k ergibt sich unmittelbar aus der Definition der Kohärenz: $A_k = AA^d$.

Um die Identität III_k zu beweisen, bedienen wir uns der Formeln I_d, 3_d und 4_d.

Man erhält nämlich

$$A^{ckck} = (A^cA^{cd})^c(A^cA^{cd})^{cd} = (A + A^{cdc})(A + A^{cdc})^d$$
$$= (A + A^{cdc})(A^d + A^{cdcd}) = AA^d + A^{cdc}A^d + AA^{cdcd} + A^{cdc}A^{cdcd}.$$

Aus 4_d folgt aber $A^{cdc}A^d = A^{cdc}A^{cdcd} = A^{cdc}$ und $AA^{cdcd} \subset AA^d$, also

$$A^{ckck} = AA^d + A^{cdc}.$$

Andererseits folgt aus I_d und 3_d:

$$A^{kckc} = [(AA^d)^c(AA^d)^{cd}]^c = AA^d + (AA^d)^{cdc}$$
$$= AA^d + (A^{cd} + A^{dcd})^c = AA^d + A^{cdc}A^{dcdc}.$$

Da aber, nach 3_d, $A^{cdc}A^{dcdc} = A^{cdc}$, erhält man

$$A^{kckc} = AA^d + A^{cdc}.$$

Die Formel III_k ist also auch bewiesen.

3. Um die Unabhängigkeit der Axiome I_k-III_k zu beweisen, nehmen wir an, der Raum C bestehe aus drei Elementen a, b, c, also: $C = (a, b, c)$. In der nachstehenden Tabelle findet man in jeder der drei Kolonnen solche Definitionen der Kohärenz aller Teilmengen von C, die allen Axiomen ausser dem einzigen oberhalb der betreffenden Kolonne markierten genügen.

	I_k	II_k	III_k
$0^k =$	0	0	0
$(a)^k =$	(a)	(a, b, c)	0
$(b)^k =$	(b)	(b, c)	0
$(c)^k =$	(c)	(c)	0
$(a, b)^k =$	0	(a, b, c)	0
$(b, c)^k =$	0	(b, c)	0
$(c, a)^k =$	0	(a, b, c)	0
$(a, b, c)^k =$	(a, b, c)	(a, b, c)	0

3. Die allgemeinen Eigenschaften der Kohärenz

1. In diesem Teil beweise ich folgende Sätze:

1_k: $\quad Aus\ A \subset B\ folgt\ A^k \subset B^k$;

2_k: $\quad (AB)^k \subset A^kB^k$;

3_k: $\quad 0^k = 0$;

4_k: $\quad C^k = C$;

5_k: $\quad A^{ck} \subset A^{kc}$;

6_k: $\quad (A - B)^k \subset A^k - B^k$;

7_k: $\quad A^{kckk} = A^{ckckck}$;

8_k: $\quad A^{kkck} = A^{kckckc}$.

2. Beweise der Sätze 1_k-8_k:

(1) Wenn man $A \subset B$ voraussetzt, bekommt man aus I_k: $A^k+B^k \subset B^k$, also $A^k \subset B^k$.

(2) Aus den evidenten Relationen $AB \subset A$ und $AB \subset B$ bekommt man, nach 1_k, $(AB)^k \subset A^k$, $(AB)^k \subset B^k$, und daher, auch $(AB)^k \subset A^kB^k$.

(3) Die Formel 3_k folgt unmittelbar aus dem Axiome II_k.

(4) Setzt man in I_k $B=C$, so bekommt man $A^k+C^k \subset C^k$; es ist also die Kohärenz jeder beliebigen Menge in der Kohärenz von C enthalten. Insbesondere hat man $\{C^{kc}\}^k \subset C^k$, also $C^{kck}C^{kc}=0$, und nach II_k: $C^{kck}=0$, also:

$$(\alpha) \qquad C^{kckc} = C.$$

Aus III_k erhält man aber

$$C^{ckck} = C^{kckc} \text{ oder } 0^{kck} = C^{kckc}$$

und nach 3_k:

$$0^{ck} = C^k = C^{kckc}$$

und schliesslich wegen (α):

$$C^k = C.$$

(5) Aus II_k folgt $A^{ck} \subset A^c$, andererseits aber $A^c \subset A^{kc}$, also $A^{ck} \subset A^{kc}$.

(6) Aus 2_k und 5_k folgert man jetzt:

$$(AB^c)^k \subset A^cB^{ck} \subset A^cA^{kc},$$

das heisst den Satz 6_k.

(7) Der Satz 7_k ergibt sich aus III_k:

$$A^{ckckck} = (A^{ckck})^{ck} = A^{kckcck} = A^{kckk}.$$

(8) Aus demselben Axiom bekommt man den Satz 8_k:

$$A^{kekekc} = (A^{kc})^{kekc} = A^{kcckck} = A^{kkck}.$$

4. Die wiederholte Anwendung der Operationen A^k und A^c

1. In diesem Teil beantworte ich zwei folgende Fragen:

(1) Welche verschiedene Mengen erhält man, wenn man auf eine beliebige Menge A eine beliebige endliche Anzahl von Operationen A^k und A^c in beliebiger Reihenfolge anwendet?

(2) Welche Relationen der Inklusion bestehen zwischen den so erhaltenen Mengen?

2. Das Symbol $A \rightarrow B$ im Folgenden bedeuten, dass $A \subset B$.

Wir betrachten jetzt zwei folgende unendliche Tabellen und beweisen die in ihnen enthaltenen Relationen.

$$\begin{array}{ccccccc}
A & \rightarrow & A^{ckc} & \rightarrow & A^{ckkc} & \rightarrow & A^{ckkkc} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{k} & \rightarrow & A^{ckck} & \rightarrow & A^{ckkck} & \rightarrow & A^{ckkkck} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{kk} & \rightarrow & A^{ckckk} & \rightarrow & A^{ckkckk} & \rightarrow & A^{ckkkckk} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{kkk} & \rightarrow & A^{ckckkk} & \rightarrow & A^{ckkckkk} & \rightarrow & A^{ckkkckkk} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot
\end{array}$$

$$\begin{array}{ccccccc}
A^{c} & \rightarrow & A^{kc} & \rightarrow & A^{kkc} & \rightarrow & A^{kkkc} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{ck} & \rightarrow & A^{kck} & \rightarrow & A^{kkck} & \rightarrow & A^{kkkck} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{ckk} & \rightarrow & A^{kckk} & \rightarrow & A^{kkckk} & \rightarrow & A^{kkkckk} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{ckkk} & \rightarrow & A^{kckkk} & \rightarrow & A^{kkckkk} & \rightarrow & A^{kkkckkk} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot
\end{array}$$

Da die zweite Tabelle aus der ersten durch Vertauschen der Mengen A und A^c ensteht, genügt es die in der oberen Tabelle enthaltenen Relationen zu beweisen.

3. Das Bildungsgesetz der Tabelle ist unmittelbar ersichtlich. Man setze nämlich

$$A^{11} = A, \quad A^{1(n+1)} = A^{1nckc} \quad \text{und} \quad A^{(m+1)n} = A^{mnk}.$$

Dann kann die obere Tabelle in folgender Weise dargestellt werden:

$$\begin{array}{ccccccc}
A^{11} & \rightarrow & A^{12} & \rightarrow & A^{13} & \rightarrow & A^{14} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{21} & \rightarrow & A^{22} & \rightarrow & A^{23} & \rightarrow & A^{24} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{31} & \rightarrow & A^{32} & \rightarrow & A^{33} & \rightarrow & A^{34} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
A^{41} & \rightarrow & A^{42} & \rightarrow & A^{43} & \rightarrow & A^{44} \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot
\end{array}$$

Die Relationen $A^{mn} \subset A^{(m+1)n}$ folgen unmittelbar aus dem Axiom II_k.

Um die Relationen $A^{m(n+1)} \subset A^{mn}$ zu beweisen genügt es die Formel $A^{m(n+1)} = A^{mnck}$ zu beweisen weil, aus II_k, $A^{ck} \subset A^c$, also für jede Menge A die Formel $A \subset A^{ck}$ folgt.

4. **Beweis der Formel** $A^{m(n+1)} = A^{mnck}$. Wir bemerken zuerst, dass man aus der Tabelle folgende Formel bekommt:

$$A^{mn} = A^{(1n)(m1)} = A^{ck\cdots kck\cdots k} = A^{ckn-1ckm-1}, \tag{1}$$

wo in $ck \cdots kck \cdots k$ die Anzahl der Faktoren k hinter dem ersten c gleich $n-1$ und hinter dem zweiten $m-1$ ist.

Die Formel

$$A^{m(n+1)} = A^{mnck} \tag{α}$$

gilt für $m=n=1$.

Wir setzen jetzt voraus, dass sie für $n=1$ und $m=r$ gilt und beweisen, dass sie dann auch für $n=1$ und $m=r+1$ gilt.

Wir haben jetzt nach (1)

$$A^{(r+1)2} = A^{ckckr} = A^{ckckr-1k} = A^{r2k} = A^{r1ckck} = A^{r1kckc} = A^{(r+1)1ckc}$$

und die Formel (α) ist für $n=1$ und jedes natürliche m bewiesen.

Nun setzen wir voraus, dass sie auch für $n=s$ gilt, dass also

$$A^{m(s+1)} = A^{msckc} \text{ oder } A^{cksckm-1} = A^{cks-1ckm-1ckc}.$$

Jetzt erhalten wir

$$\begin{aligned} A^{m(s+2)} &= A^{cks+1ckm-1} = A^{ckksckm-1} = A^{ckccksckm-1} \\ &= A^{ckccks-1ckm-1ckc} = A^{cksckm+1ckc} = A^{m(s+1)ckc}. \end{aligned}$$

Wir haben also allgemein den Satz bewiesen:

$$A^{mn} \subset A^{st} \text{ für } s \leqq m,\ t \geqq n,$$

$$A^{mn} \subset A^{st} \text{ für } s \geqq m,\ t \leqq n.$$

5. Dass es keine weiteren Relationen zwischen den in beiden Tabellen enthaltenen Mengen bestehen, ersieht man aus dem folgenden Beispiel:

Es sei C ein linearer Raum (eine Gerade). Auf dieser Geraden wählen wir zwei fremde Strecken. Auf der ersten von ihnen wählen wir eine wohlgeordnete Menge vom Typus $(\omega+1)^{\omega+1}$ und bezeichnen diese Menge mit M. Mit N bezeichnen wir die Differenz der zweiten Strecke und einer beliebigen in ihr enthaltenen wohlgeordneten Menge vom Typus $(\omega+1)^{\omega+1}$. Die verlangte Menge A ist die Summe $A=M+N$.

Wenn man auf die so definierte Menge A alle in beiden Tabellen enthaltenen Operationen anwendet, bekommt man lauter verschiedene Mengen, zwischen denen ausser den in den Tabellen angegebenen keine anderen Relationen der Inklusion bestehen.

6. Man ersieht leicht, dass man durch wiederholte Anwendung der Formeln III_k, 7_k, 8_k und $A^{cc}=A$ jede beliebige aus A durch successive Anwendung der Operationen A^k und A^c erhaltene Menge in eine der in den Tabellen enthaltenen Mengen identisch überführen kann. Es gibt also keine von den in den Tabellen gegebenen Mengen verschiedene Menge, die aus A durch die in beliebiger Reihenfolge angewandten Operationen A^k und A^c erhalten werden kann.

Die am Anfang des §4 gestellten Fragen sind somit volkommen beantwortet.

5. Die Cantorschen Adhärenzen

1. G. Cantor nannte die Menge $A^{a_n}=A^{k^{n-1}}A^{k^n c}$* die $n$te Adhärenz von A. Es sei bemerkt, dass die $n$te Adhärenz von A keineswegs mit der ersten Adhärenz der $(n-1)$ten Adhärenz von A identisch ist. Es gelten vielmehr die Identitäten

$$A^{a_n a_1}=A^{a_n} \text{ und } A^{a_n a_m}=0 \text{ für jedes } n \text{ und } m \geqq 2.$$

2. Jede Menge A lässt sich (für jede natürliche Zahl n) auf folgende Weise in elementenfremde Mengen zerlegen:

$$(\alpha) \qquad A = A^{a_1}+A^{a_2}+\cdots+A^{a_n}+A^{k^n}.$$

Man beweist leicht, dass die Summanden obiger Summe wirklich elementenfrei Mengen bilden. Wir haben nämlich

$$A^{k^m} \subset A^{k^n} \text{ für } m>n, \text{ also : } A^{k^m}A^{k^n c}=0.$$

---

* A^{k^n} ist die $n$te Kohärenz von A.

Es ist aber

$$A^{a_m} = A^{k^{m-1}} A^{k^m c} \text{ und } A^{a_n} = A^{k^{n-1} k^n c}$$

und daher:

$$A^{a_m} A^{a_n} = 0.$$

Es ist auch evident, dass für $s \leqq n$:

$$A^{k^n} A^{a_s} = A^{k^n} A^{k^{s-1}} A^{k^s c} = 0, \text{ weil } A^{k^n} \subset A^{k^s}.$$

Die Formel (α) ist für $n=1$ giltig, denn $A^{a_1}+A^k=AA^{kc}+A^k=A$, weil A^k in A enthalten ist. Setzen wir nun voraus die Formel (α) gelte für $n=r$. Wir bekommen dann

$$A^{a_1} + A^{a_2} + \cdots + A^{a_r} + A^{k^r} = A.$$

Es ist aber

$$A^{a_{r+1}} + A^{k^{r+1}} = A^{k^r} A^{k^{r+1} c} + A^{k^{r+1}} = A^{k^r},$$

weil $A^{k^{r+1}} \subset A^{k^r}$. Man erhält also

$$A^{a_1} + A^{a_2} + \cdots + A^{a_{r+1}} + A^{k^{r+}} = A,$$

womit die sonst bekannte Formel (α) bewiesen ist.

3. Wir liefern noch einen einfachen Beweis eines von W. H. Young* aufgestellten, aber erst von L. E. J. Brouwer† richtig bewiesenen Satzes:

(β) $$A^{a_n} \subset A^{a_m d} \text{ für jedes } m < n.$$

Die Formel (β) gilt für $m=1$ und $n=2$, denn

$$A^{a_2} = A^k A^{kkc} = AA^d \left\{AA^d (AA^d)^d\right\}^c = AA^d \left\{A^c + A^{dc} + (AA^d)^{dc}\right\}$$
$$= AA^d (AA^d)^{dc}.$$

Es ist aber nach 2_d:

$$A^d (AA^d)^{dc} \subset \left\{A(AA^d)^c\right\}^d = \left\{A(A^c + A^{dc})\right\}^d (= AA^{dc})^d,$$

also:

$$A^{a_2} \subset A(AA^{dc})^d.$$

Andererseits bekommt man

$$A^{a_1 d} = (AA^{kc})^d = \left\{A(AA^d)^c\right\}^d = (AA^{dc})^d.$$

Die Formel $A^{a_2} \subset A^{a_1 d}$ gilt also für jede Menge, also auch für die Menge $A^{k^{n-2}}$ und man bekommt jetzt

$$A^{k^{n-2} a_2} \subset A^{k^{n-2} a_1 d}.$$

---

* W. H. Young, Quarterly Journal of Mathematics, vol. 35 (1903).

† L. E. J. Brouwer, Amsterdam Academy Proceedings, vol. 18 (1915).

Wir haben aber

$$A^{k^{n-2}a_2} = A^{k^{n-2}ka_1} = A^{k^{n-1}a_1} = A^{a_n}$$

und

$$A^{k^{n-2}a_1 d} = A^{a_{n-1}d}, \text{ also } A^{a_n} \subset A^{a_{n-1}d} \qquad (n = 1,2, \cdots).$$

Jede Adhärenz ist also in der Ableitung der im Konstruktionsmodus unmittelbar vorangehenden Adhärenz enthalten.

Aus der letzten Formel erhält man aber nach 1_d und III_d:

$$A^{a_{n+1}d} \subset A^{a_n dd} \subset A^{a_n d},$$

und daher

$$A^{a_n} \subset A^{a_{n-1}d} \subset A^{a_{n-1}d}.$$

Durch dieselbe Schlussweise bekommt man endlich

$$A^{a_n} \subset A^{a_m d} \text{ für jedes } m < n.$$

LEMBERG, POLAND

ON THE IRREGULAR CASES OF THE LINEAR ORDINARY DIFFERENCE EQUATION*

BY

C. RAYMOND ADAMS

Introduction. In the analytic theory of the linear ordinary difference equation

$$\sum_{i=0}^{n} a_i(x)g(x+n-i) = 0 \tag{1}$$

whose coefficient functions are rational and expressible in the form

$$a_i(x) = x^m(a_{i0} + a_{i1}x^{-1} + a_{i2}x^{-2} + \cdots) \qquad \text{for } |x| > R \tag{2}$$
$$(i = 0,1,2, \cdots, n)$$

and whose *characteristic equation* is

$$a_{00}\rho^n + a_{10}\rho^{n-1} + \cdots + a_{n-1,0}\rho + a_{n0} = 0, \tag{3}$$

the assumption is commonly made that the roots of (3) are finite, distinct, and different from zero. This narrowly restrictive hypothesis characterizes what we call the *regular case*. But little attention has yet been devoted to the more general problem of the *irregular cases*, in which the roots fail to satisfy this hypothesis.

Barnes† in 1905 studied most of the irregular cases of the equation of second order with coefficients linear in x, employing the Laplace transformation.

In 1910 Horn‡ gave formal series solutions (divergent) and a theorem setting forth the existence of solutions whose asymptotic forms for large real positive values of x are given by the formal series, when the only departure from regularity is the vanishing of one root.

In papers read (by title) before the Society in 1913 Batchelder§ found formal series solutions in all possible irregular cases for the equation of second

---

* Presented to the Society, December 28, 1926; received by the editors in July, 1927.

† *On the homogeneous linear difference equation of the second order with linear coefficients*, Messenger of Mathematics, vol. 34 (1905), pp. 52–71.

‡ *Über das Verhalten der Integrale linearer Differenzen- und Differentialgleichungen für grosse Werte der Veränderlichen*, Crelle's Journal, vol. 138 (1910), pp. 159–191, in particular p. 191.

§ *The divergent series satisfying linear difference equations of the second order* and *The hypergeometric difference equation*, Bulletin of the American Mathematical Society, vol. 19 (1912–13), pp. 498, 500–502.

order and sought to extend the theory as developed by Birkhoff* for the regular case to the irregular cases when the coefficient functions are linear in x. Batchelder's results are contained in a Dissertation† deposited in the Harvard Library in 1916; they are also incorporated in *An Introduction to Linear Difference Equations*, presently to be published with the aid of the National Research Council.

Perron‡ in 1917 derived certain properties of solutions of the equation of the second order whose characteristic equation has equal roots.

The most searching investigation yet made of the irregular cases for the equation of nth order is by Galbrun.§ In 1921 he employed the Laplace transformation, as he had done earlier in his study of the regular case,‖ in developing the theory of equation (1) when two of the roots of (3) are equal.

Nörlund, in his *Differenzenrechnung*,¶ does little more than mention the problem of the irregular cases.

The purposes of this paper are first to obtain formal power series solutions of the equation (1) in the irregular cases and secondly to adapt the general methods employed by Birkhoff in the regular case to the problem of finding analytic solutions which shall be asymptotically represented by the formal series. In §1 formal series solutions are found for an extensive class of irregular cases. §§2, 3, and 4 are devoted to a class of irregular cases which have a theory much like that of the regular case: in §2 is proved the existence of two sets of solutions resembling the principal solutions of the regular case; in §3 the periodic functions defined by these two sets of solutions are considered briefly; and in §4 the asymptotic behavior of these two sets of solutions in the entire plane is examined. A second class of irregular cases is treated in §§5, 6, 7, and 8: §5 is concerned with formal series in the case in which (3) has an n-fold root not zero; in §6 certain existence theorems are proved in the same case; in §7 the results and implications of §§5 and 6 for $n=2$ are given special attention; and in §8 are pointed out certain

---

* *General theory of linear difference equations*, these Transactions, vol. 12 (1911), pp. 243–284.

† *The hypergeometric difference equation.*

‡ *Über lineare Differenzengleichungen zweiter Ordnung deren charakteristische Gleichung zwei gleiche Wurzeln hat*, Sitzungsberichte der Heidelberger Akademie der Wissenschaften (mathematisch-naturwissenschaftliche Klasse), vol. 8A (1917), No. 17, 18pp.

§ *Sur certaines solutions exceptionnelles d'une équation linéaire aux différences finies*, Bulletin de la Société Mathématique de France, vol. 49 (1921), pp. 206–241.

‖ *Sur la représentation des solutions d'une équation linéaire aux différences finies pour les grandes valeurs de la variable*, Acta Mathematica, vol. 36 (1913), pp. 1–68.

¶ Berlin, Springer, 1924, pp. 339–342.

results which can be inferred for a large class of irregular cases from the study of a particular case made in §§5, 6, and 7. Throughout the paper an attempt is made to keep the notation and arrangement parallel to Birkhoff's, and it is assumed that his treatment of the regular case is familiar to the reader.

We make the hypothesis that the equation (1) is *irreducible*; that is, that it is not satisfied by any analytic function of x which is a solution of a difference equation of order less than n with rational coefficients.

1. **Formal series solutions.** We shall set forth here only a statement of the types of series that formally satisfy (1) in various irregular cases; that these series do satisfy (1) can be verified directly by substitution, although the labor involved is not inconsiderable. The cases in which the only irregularities are multiple roots finite and different from zero are grouped for consideration in

CLASS 1 ($a_{00} \neq 0$, $a_{n0} \neq 0$). Corresponding to a simple root ρ_1 there is one formal series of the "regular" type:*

$$s(x) = \rho_1{}^x x^r P(x^{-1}). \tag{4}$$

Corresponding to a root ρ_1 of multiplicity $m > 1$, and on the assumption that ρ_1 is not a root of the *secondary equation*

$$a_{01}\rho^n + a_{11}\rho^{n-1} + \cdots + a_{n-1,1}\rho + a_{n1} = 0, \tag{5}$$

there are m series of the following type:*

$$s(x) = \rho_1{}^x e^{L^{(m)}(x)} x^r P(x^{-1/m}), \tag{6}$$

where

$$L^{(m)}(x) = \gamma^{(m-1)} x^{(m-1)/m} + \gamma^{(m-2)} x^{(m-2)/m} + \cdots + \gamma' x^{1/m},$$

the constants $\gamma^{(m-1)}$, $\gamma^{(m-2)}$, etc., being different in the different series. The m values of $\gamma^{(m-1)}$ are the m determinations of the mth root of a constant not zero times the left-hand member of (5) after ρ_1 has been substituted for ρ,

---

* r is a constant, in this and subsequent types. $P(x^{-1})$ stands for a power series in x^{-1}, namely $1+s'x^{-1}+s''x^{-2}+\cdots$; the constant term may be taken as 1, since we are dealing with the homogeneous difference equation. Similarly, $P(x^{-1/m})$ is used to represent a power series in $x^{-1/m}$, the first term being 1. The γ's, wherever they occur, are constants. In each type of series the constants are calculated (by formal substitution of the series in (1)) in the order in which they appear in the series as written; for example, in (6) the constants after ρ_1 are calculated in the order $\gamma^{(m-1)}, \gamma^{(m-2)}, \cdots, \gamma', r, s', s'', \cdots$. The constants thus calculated are uniquely determined for each series of the type.

and hence are different from zero. It is easily seen that these m series are the m determinations of a single series.*

Corresponding to a root ρ_1 of multiplicity $m>2$ which is a *simple* root of the secondary equation (5), there is a set of m series as follows:

1 series, $\quad s(x) = \rho_1^x x^r P(x^{-1/(m-1)})$;

$m-1$ series, $\quad s(x) = \rho_1^x e^{L^{(m-1)}(x)} x^r P(x^{-1/(m-1)})$.

In the case $m=2$, these reduce to "regular" series like (4), but in order that the equation (1) be satisfied by two such series, further conditions must be fulfilled.

When ρ_1 is an m-fold root of the characteristic equation and a multiple root of the secondary equation, additional complications enter the problem of obtaining formal solutions. In general under these conditions we have been able to obtain only part of the full quota of m series; in particular cases the full number has been found, but the facts that the cases are particular and that the statements of conditions are long and involved lead us to omit their description here. Two points may deserve mention: (i) that the presence of ρ_1 as an m_1-fold root of the secondary equation tends to reduce to $m-m_1$ the index of the root of x^{-1} according to powers of which the series proceed; and (ii) that if m_1 is sufficiently large, the question of whether ρ_1 satisfies the subsequent equations

$$a_{0j}\rho^n + a_{1j}\rho^{n-1} + \cdots + a_{n-1,j}\rho + a_{nj} = 0 \qquad (j = 2,3,\cdots) \tag{7}$$

becomes of importance.

We desire to call attention to only one further case. If ρ_1 is an m-fold root of (3), a root of multiplicity $\geqq m$ of (5), a root of multiplicity $\geqq m-1$ of (7) for $j=2$, a root of multiplicity $\geqq m-2$ of (7) for $j=3,\cdots$, a root of multiplicity $\geqq 2$ of (7) for $j=m-1$, and is *not* a root of (7) for $j=m$, then the equation (1) is satisfied by m series of the regular type (4). In the event of all these hypotheses except the last being satisfied, and if ρ_1 satisfies all the equations (7) for $j\geqq m$ (as it would if, for example, the coefficient functions (2) were polynomials of degree $\leqq m-1$), an analytic solution of the equation (1) is ρ_1^x ; but this function satisfies a difference equation of first order with rational coefficients, and (1) is therefore reducible.

Secondly we consider the cases in which some or all of the irregularities are due to the presence of zero or infinite roots or of both; these cases we group in

* Cf. §5, in which this fact is pointed out more clearly in a particular case.

CLASS 2 (one or both of $a_{00}, a_{n0}=0$). Let us denote by a_{i,j_i} the first non-zero coefficient in $a_i(x)$ $(i=0, 1, \cdots, n)$, and choosing i- and j-axes, plot the points (i, j_i) as in Figure 1. Construct a broken line L, convex upward, such that both ends of each segment of the line are points of the set (i, j_i) and such that all points of the set lie upon or beneath the line. This is the form that would be assumed by an elastic string if pegs were inserted at the points (i, j_i), the ends of the string fastened one at $(0, j_0)$ and the other at (n, j_n), and the string allowed to contract from above upon the pegs. At least one of the points (i, j_i) will clearly be situated on the i-axis; otherwise a power of x might be suppressed in the entire equation (1).

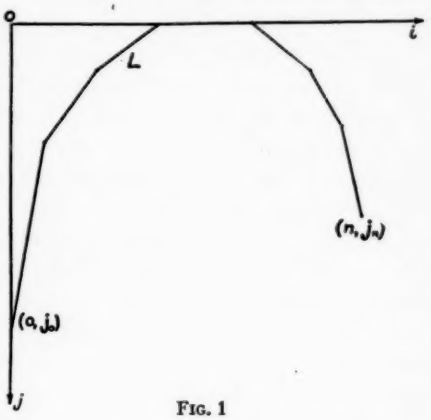

FIG. 1

In this class of irregular cases the so-called "characteristic equation" (3) *is by no means completely characteristic of the difference equation* (1). In fact we would rather say that (3) *is replaced by several characteristic equations, one associated with each segment of* L. The degree of the characteristic equation associated with any segment of L is 1 less than the number of points (i, j_i) that lie on or beneath that segment (inclusive of its end points). The coefficients of this characteristic equation are the a_{i,j_i} corresponding to points (i, j_i) actually *on* that segment of L; the coefficient corresponding to a point (i, j_i) *beneath* the segment is zero. Evidently the sum of the degrees of these several characteristic equations is n. If one of the segments is horizontal, the characteristic equation associated with that segment (this may be obtained from (3) by suppressing the zero and infinite roots) picks out its quota of formal solutions precisely as (3) distinguishes n solutions in any case of Class 1. As for a segment not horizontal, let the slope of any such segment be $-\mu$, a rational number different from zero. The transformation*

$$f(x) = x^{\mu x}e^{-\mu x}g(x) \tag{8}$$

* In making this transformation one should employ the expansion

$$\left(1+\frac{i}{x}\right)^{\mu x} = e^{\mu x \log[1+i/x]} = e^{i\mu}\left(1-\frac{i^2\mu}{2x}+\frac{i^2\mu(3i\mu+8)}{24x^2}-\cdots\right).$$

The factor $e^{-\mu x}$ is inserted in (8) merely for convenience in annulling the $e^{i\mu}$ here.

then changes (1) into a new equation of exactly the same type except in the respect that if $-\mu$ is not an integer but a fraction, q/p in lowest terms (p positive), some of the coefficient functions will be of the form $x^{-s/p}A(x)$, where s is a positive integer and $A(x)$ a power series in x^{-1}. The effect of (8) upon the points (i, j_i) is to relocate them in such a way that each segment of the new broken line L' that "roofs them over" has a slope μ greater than that of the corresponding segment of L. Thus, in particular, the segment of L whose slope is $-\mu$ becomes a horizontal segment of L'.

This analysis makes it clear that if $-\mu$ is an integer, the state of affairs with respect to the segment of L having this slope is wholly similar to the situation relative to a horizontal segment and hence analogous to the cases of Class 1. The formal series associated with the segment of slope $-\mu$ are like those enumerated above under Class 1 except for the additional factors $x^{\mu x}e^{-\mu x}$ precedineg the power series itself. The equation which plays the rôle of (5) here is th equation whose coefficients are the a_{ij} corresponding to points (i, j) which if plotted in Figure 1 would lie upon a line parallel to this segment of slope $-\mu$ and one unit vertically below it. The tertiary equation for a segment of slope $-\mu$ is one whose coefficients a_{ij} are those whose corresponding points (i, j) if plotted would lie upon a line two units vertically below the segment, and so on for the subsequent equations.

If on the other hand $-\mu$ is a fraction, q/p in lowest terms, there corresponds to each simple root ρ_1 of the characteristic equation for this segment of L a formal solution of the type

$$s(x) = x^{\mu x}e^{-\mu x}\rho_1{}^x e^{L^{(p)}(x)}x^r P(x^{-1/p}). \tag{9}$$

It is noteworthy that in this case $\gamma^{(p-1)}$, and in addition any or all of the subsequent γ's in the exponent of e, may vanish. In fact if all the points (i, j_i) not on the segment of slope $-\mu$ are situated on or beneath a line parallel to this segment and one unit vertically below it, the γ's in (9) will all be zero.

When, for $-\mu = q/p$, the characteristic equation for this segment of L has a root ρ_1 of multiplicity $m > 1$, the situation is more complicated. Let the segment of slope q/p be prolonged in both directions to form a line l. Of the points (i, j) corresponding to non-zero coefficients a_{ij} but not situated on this segment there will be one or more whose distance below l, measured vertically, is least; let that distance be t/p and draw a line l' parallel to l through this point (or these points). Let the constants a_{ij} corresponding to points (i, j) on l' be used as coefficients in an algebraic equation of degree n; this is the *secondary equation for the segment of slope* q/p. If the secondary equation is not satisfied by ρ_1 and if t/m is 1 or a submultiple of 1, then

corresponding to the root ρ_1 there are m series of type (9) with p replaced by w, where $w=t/mp$. When the secondary equation has ρ_1 as a root, the situation is analogous to that described in the corresponding case under Class 1 above.

CASES IN WHICH THE THEORY RESEMBLES CLOSELY THAT OF THE REGULAR CASE

2. **Existence theorems.** In certain of the irregular cases the theory of equation (1) is much like that of the regular case; these cases fall under the following classes.

CLASS 2a. *The slope of each segment of L is an integer and the characteristic equation associated with each segment has only simple roots.*

CLASS 2b. *The slopes of some or all of the segments of L are fractional; the characteristic equation associated with a segment of L whose slope is an integer has only simple roots; the characteristic equation associated with a segment of L whose slope is fractional has only simple roots and either* (a) *no two of these roots are of equal absolute value, or* (b) *if two or more of these roots have the same absolute value, no exponential factor $e^{L^{(p)}(z)}$ occurs in the formal series corresponding to them* (conditions under which the situation (b) would obtain are described in §1).

We point out briefly in this section the dissimilarities between the existence theorems in the cases of Class 2a and in the regular case; the discussion requires only slight modifications, chiefly in respect to the formulas, to adapt it to the cases of Class 2b.

In each case of Class 2a the equation (1) possesses n formal series solutions which we denote by

$$(10)\quad s_i(x) = x^{\mu_i x} e^{-\mu_i x} \rho_i^x x^{r_i}(1 + s_i' x^{-1} + s_i'' x^{-2} + \cdots) \quad (i = 1,2, \cdots, n).$$

All the μ_i for the set of solutions associated with a particular segment of L then have the same value. To gain the simplicity of the matrix notation and to make our work parallel that of Birkhoff, we write our single equation (1) of the nth order in the form of a system of n linear equations of the first order; the n formal solutions (10) then provide us with n sets of formal solutions for the system.* These sets we arrange in a matrix, of which each set constitutes a column; the elements of the first row are the n series (10). The order of the columns in the matrix is of importance. Let them be arranged first according to descending values of μ_i; secondly, let those for which μ_i is the same be ordered according to descending values of $|\rho_i|$.

* The details of this frequently employed device are shown fully in a particular case in §5.

Then $(x/e)^{\mu_i}\rho_i$ plays the rôle that ρ_i does in the regular case, and since we have

$$\left|\left(\frac{x}{e}\right)^{\mu_i}\rho_i\right| \geqq \left|\left(\frac{x}{e}\right)^{\mu_j}\rho_j\right|, \qquad i > j, \tag{11}$$

for large values of $|x|$ when μ_i is $>\mu_j$ and for all values of x when μ_i is $=\mu_j$, no difficulty is experienced in establishing the existence of determinant limits,* of solutions associated with them, and of intermediate solutions.

In seeking solutions by the aid of contour integrals, however, we must in some instances make a different choice of the λ_{jk} in Birkhoff's formula (40) in order to insure that $g_{1k}(x)$ be asymptotically represented by $s_{1k}(x)$ in as large a region as possible. For x above $A_1\infty$ we have, as in the regular case,†

$$\begin{aligned} g_{1k}(x) \sim\ & s_{11}(x)e^{2\pi\lambda_{1k}(-1)^{1/2}x}q_{1k}(x) + \cdots \\ & + s_{1,k-1}(x)e^{2\pi\lambda_{k-1,k}(-1)^{1/2}x}q_{k-1,k}(x) + s_{1k}(x), \end{aligned} \tag{12}$$

and we desire the last term on the right to be the dominant one in the left half-plane. The governing influences in the terms are the exponential factors, which are

$$\left(\frac{x}{e}\right)^{\mu_1 x}\rho_1^{x}e^{2\pi\lambda_{1k}(-1)^{1/2}x}, \cdots, \left(\frac{x}{e}\right)^{\mu_{k-1}x}\rho_{k-1}^{x}e^{2\pi\lambda_{k-1,k}(-1)^{1/2}x}, \left(\frac{x}{e}\right)^{\mu_k x}\rho_k^{x}.$$

Writing all these as exponentials to the base e and dividing through by the last, we may express the exponents in the following form:

$$\begin{aligned} & 2\pi(-1)^{1/2}\left(\lambda_{1k} - \frac{\mu_1 - \mu_k + \log\rho_k - \log\rho_1 - (\mu_1 - \mu_k)\log x}{2\pi(-1)^{1/2}}\right)x, \cdots, \\ & 2\pi(-1)^{1/2}\left(\lambda_{k-1,k} - \frac{\mu_{k-1} - \mu_k + \log\rho_k - \log\rho_{k-1} - (\mu_{k-1} - \mu_k)\log x}{2\pi(-1)^{1/2}}\right)x,\ 0. \end{aligned} \tag{13}$$

We propose to select the λ's so that these exponents, save the last, will have real parts that become negatively infinite as x becomes infinite in the second quadrant. If the $\mu_i (i = 1, 2, \cdots, k)$ in (13) are all equal, these exponents are identical with the corresponding exponents in the regular case. This makes it clear that the λ's may be chosen so that those solutions on the left,

---

* We employ the terminology of Birkhoff, loc. cit.

† It should be observed that, to be consistent with the definition of asymptotic representation of a function $g(x)$ (cf. Birkhoff, loc. cit., p. 248), the relation (12) should, until the question of dominance is settled, be interpreted as meaning "$g_{1k}(x)$ is the sum of k functions which have for asymptotic forms the several terms on the right." For simplicity, however, we shall continue to write such relations in the form of (12).

obtained by contour integrals, which are associated with the segment of L farthest to the left will have the same properties as do all the principal solutions in the regular case. When μ_i is thus equal to μ_k, we choose λ_{ik} as the least integer exceeding*

$$\frac{\arg \rho_k - \arg \rho_i}{2\pi}. \tag{14}$$

When μ_i is greater than μ_k, the ith exponent in (13) clearly has a real part that becomes negatively infinite as x recedes to infinity in the sector $\pi/2+\epsilon \leqq \arg x < \pi$, ϵ being an arbitrarily small positive number, *whatever choice of* λ_{ik} *be made*; this is owing to the presence of $\log x$ within the parentheses. If, however, we choose λ_{ik} to be any integer (*and we take it to be the least integer*) greater than

$$\alpha_{ik} = \frac{\arg \rho_k - \arg \rho_i - (\mu_i - \mu_k)\pi/2}{2\pi} \tag{15}$$

the asymptotic form of $g_{1k}(x)$ will be given by $s_{1k}(x)$ for x above $A_1\infty$ and on or to the left of the imaginary axis, or likewise on or to the left of any parallel to it.

For x in the strip bounded by $A_1\infty$ and $B_1\infty$ the asymptotic form of $g_{1k}(x)$ is given by $s_{1k}(x)$ as in the regular case, by virtue of the ordering of the formal series according to μ's and ρ's.

When x is below $B_1\infty$ we have

$$\begin{aligned} g_{1k}(x) \sim s_{11}(x)e^{2\pi(\lambda_{1k}-1)(-1)^{1/2}z}q_{1k}(x) + \cdots \\ + s_{k-1,k}(x)e^{2\pi(\lambda_{k-1,k}-1)(-1)^{1/2}z}q_{k-1,k}(x) + s_{1k}(x). \end{aligned} \tag{16}$$

The dominance depends upon the real part of the exponents

$$\begin{aligned} &2\pi(-1)^{1/2}\left(\lambda_{1k} - 1 - \frac{\mu_1 - \mu_k + \log \rho_k - \log \rho_1 - (\mu_1 - \mu_k)\log x}{2\pi(-1)^{1/2}}\right)x, \cdots, \\ &2\pi(-1)^{1/2}\left(\lambda_{k-1,k} - 1 - \frac{\mu_{k-1} - \mu_k + \log \rho_k - \log \rho_{k-1} - (\mu_{k-1} - \mu_k)\log x}{2\pi(-1)^{1/2}}\right)x, 0. \end{aligned} \tag{17}$$

For $\mu_i = \mu_k (i = 1, 2, \cdots, k-1)$ the asymptotic form of $g_{1k}(x)$ in any sector $\pi < \arg x \leqq 3\pi/2 - \epsilon$ for which ϵ is positive is $s_{1k}(x)$. If none of the quantities

* This choice is the same as Birkhoff's in the regular case when the quantity (14) is not an integer; if (14) is an integer, our λ_{ik} exceeds his by 1. We make this choice in order to insure that $g_{1k}(x)$ always be represented asymptotically by $s_{1k}(x)$ in the direction of the positive axis of imaginaries; with Birkhoff's choice and (14) an integer for one or more values of i, the asymptotic form of $g_{1k}(x)$ in that direction is given by the sum of two or more terms of (12), one of which is the last.

(14) is an integer, we have $g_{1k}(x) \sim s_{1k}(x)$ for $\pi < \arg x \leqq 3\pi/2$; if on the other hand one or more of the quantities (14) is an integer, the asymptotic form of $g_{1k}(x)$ in the direction of the negative axis of imaginaries is given by the sum of two or more terms of (16), one of which is the last.

When μ_i is greater than μ_k for some or all of the values of $i(=1, 2, \cdots, k-1)$, the ith exponent in (17) has a real part that becomes negatively infinite as x recedes to infinity in the sector $\pi < \arg x \leqq 3\pi/2 - \epsilon$, due to the presence of the term $\log x$, so that we have $g_{1k}(x) \sim s_{1k}(x)$ in that sector. In the direction of the negative axis of imaginaries, however, the asymptotic form of $g_{1k}(x)$ is given by $s_{1k}(x)$ only when all of the following conditions are fulfilled: (a) $\mu_i - \mu_k$ is $\leqq 1$ for $i=1, 2, \cdots, k-1$; (b) when $\mu_i - \mu_k = 1$, $\lambda_{ik} - \alpha_{ik}$ is $< \frac{1}{2}$; (c) when $\mu_i - \mu_k$ is 0, the quantity (14) is not an integer. If among these conditions (c) alone fails to be satisfied, the asymptotic form of $s_{1k}(x)$ in the direction of the negative axis of imaginaries is given by the sum of two or more terms of (16); one of these terms is the last, while the others correspond to values of i for which $\mu_i = \mu_k$ and (14) is an integer. If either or both of conditions (a) and (b) fail, the asymptotic form of $s_{1k}(x)$ in the direction in question is given by the term (or sum of terms) of (16) corresponding to the value (or values) of i for which

$$\lambda_{ik} - 1 - \alpha_{ik} + \frac{\mu_i - \mu_k}{2}$$

is largest.

The functions $g_{ij}(x)$ are analytic except for poles throughout the entire finite plane; we denote the matrix $(g_{ij}(x))$ by $G(x)$.

The freedom that we have in the choice of λ_{ik} when μ_i is $> \mu_k$ makes it clear that the solutions on the left associated with segments of L other than that farthest to the left are not in general characterized uniquely by the properties we have proved for them. Exceptions can occur only when the solutions in question are associated with the second segment from the left and when the slope of that segment is only 1 greater than the slope of the first segment. We therefore hesitate to apply to these solutions the term "principal solutions" except in the case in which L consists of but a single segment; this case is, however, essentially regular.

There exists a similar set of solutions "on the right," in obtaining which we choose $\lambda_{ik} (i > k)$ as the least integer exceeding α_{ik} (cf. (15); the fact that in α_{ik}, i is now $> k$ should not be overlooked). From the relation $\alpha_{ik} = -\alpha_{ki}$ it follows that $\lambda_{ki} = 2 - \lambda_{ik}$ or $1 - \lambda_{ik}$ according as α_{ik} is or is not an integer. In this set of solutions only those associated with the segment of L farthest to the right can in general be said to be characterized by the properties we prove them to possess. We denote the matrix of these functions by $H(x)$;

its elements are analytic except for poles throughout the finite plane.

3. **Periodic functions.** We shall examine briefly the matrix of periodic functions $P(x)$ defined by

$$G(x) = H(x)P(x) ; \tag{18}$$

that is,

$$P(x) = H^{-1}(x)G(x).$$

The elements of $H^{-1}(x)$ and $G(x)$ are analytic except for poles over the entire finite plane. Hence in any period strip the elements $p_{ij}(x)$ of $P(x)$ are analytic except for poles throughout the finite portion of the strip. We set $z = e^{2\pi(-1)^{1/2}x}$ and regard $p_{ij}(x)$ as a function of z; the upper end of the strip corresponds to $z=0$, the lower end to $z=\infty$.

At the upper end of the strip we have

$$P(x) \sim S^{-1}(x)S(x),$$

where $S(x)$ is the matrix $(s_{ij}(x))$ and $s_{ij}(x)$ is $s_j(x+i-1)$; this gives us the asymptotic relation

$$p_{ij}(x) \sim \left(\frac{x}{e}\right)^{(\mu_j-\mu_i)x}\left(\frac{\rho_j}{\rho_i}\right)^{x} x^{r_j-r_i}\delta_{ij}, \qquad \delta_{ij} = \begin{cases} 1, & i=j; \\ 0, & i \neq j. \end{cases}$$

It follows that for x sufficiently high up in the strip we have

$$p_{ij}(x) = e^{2\pi\lambda_{ij}(-1)^{1/2}x}(c_{ij} + c_{ij}' e^{2\pi(-1)^{1/2}x} + \cdots), \tag{19}$$

where λ_{ii} is defined to be zero and c_{ii} is 1; the series in parentheses is convergent in the vicinity of $z=0$.

At the lower end of the strip we have a situation which in form, though not in substance, is much more complicated than the one we meet in the regular case. The asymptotic form of $G(x)$ is not in general $S(x)$, nor is the asymptotic form of $H^{-1}(x)$ given by the product of $S^{-1}(x)$ and a simple matrix. An examination of $p_{ij}(x)$ readily shows that it is analytic or becomes infinite to only a finite order at $z=\infty$, but the determination of that order explicitly in terms of the μ's and ρ's is not feasible.* We conclude that *the*

* Such a determination might be made feasible by reassigning the values of the λ's in (12) when μ_1 is $>\mu_k$ [and in the similar expression for the form of $h_{1k}(x)$ in the first quadrant above $A_1\infty$ when μ_n is $<\mu_k$] so as to insure that $g_{1k}(x)$ be given asymptotically, as x becomes infinite along a parallel to the negative axis of imaginaries, by, for example,

$$s_{11}(x)e^{X}q_{1k}(x) \qquad [s_{1n}(x)e^{Y}q_{1k}'(x)],$$

where

$$X = 2\pi(\lambda_{1k}-1)(-1)^{1/2}x, \quad Y = 2\pi(\lambda_{nk}-1)(-1)^{1/2}x,$$

and in addition by imposing the restriction that for all values of k associated with the first segment of L on the left [last segment of L on the right] the expression (14) be not an integer. This restriction, if we take for $\arg \rho_i (i=1, 2, \cdots, n)$ the principal values of the arguments, means that no two of the roots of the characteristic equation associated with the first segment of L on the left [last segment of L on the right] may have the same argument.

elements $p_{ij}(x)$ are rational functions of $e^{2\pi(-1)^{1/2}x}$. Let the order of the pole of $p_{ij}(x)$ at ∞ be denoted by ν_{ij}; then for values of x sufficiently low in the strip we have

$$p_{ij}(x) = e^{2\pi\nu_{ij}(-1)^{1/2}x}(d_{ij} + d'_{ij}e^{-2\pi(-1)^{1/2}x} + \cdots).$$

It may be remarked that if the coefficient functions of our system of equations have no finite poles, as is the case when $a_0(x)$ reduces to a constant, then the elements of $G(x)$ are analytic without exception over the finite plane and the series in parentheses in (19) are polynomials. Similarly, if the inverse of the matrix of these coefficient functions has no finite poles, and it will have none if $a_n(x)$ is a constant, the elements of $H(x)$ have no poles in the finite plane and $P^{-1}(x)$ is a matrix of functions each of which is a polynomial in z multiplied by some power of z. Except in respect to the formulas this section needs practically no modification to adapt it to the cases of Class 2b, §2.

4. **The solutions $G(x)$ and $H(x)$ in the entire plane.** We are now in a position to examine the asymptotic form of $g_{ij}(x)$ (or $h_{ij}(x)$) in the entire plane. The equation (18) gives us

$$g_{ij}(x) = h_{i1}(x)p_{1j}(x) + \cdots + h_{in}(x)p_{nj}(x).$$

Along a ray from the origin in the first quadrant above the axis of reals we have

$$p_{ij}(x) \sim e^{2\pi\lambda_{ij}(-1)^{1/2}x}c_{ij},$$

and therefore

$$g_{ij}(x) \sim s_{i1}(x)e^{2\pi\lambda_{1j}(-1)^{1/2}x}c_{1j} + \cdots + s_{in}(x)e^{2\pi\lambda_{nj}(-1)^{1/2}x}c_{nj}. \tag{20}$$

The question of which term dominates turns upon the relative magnitude of the absolute values of

$$\left(\frac{x}{e}\right)^{\mu_1 x}\rho_1{}^x e^{2\pi\lambda_{1j}(-1)^{1/2}x}, \cdots, \left(\frac{x}{e}\right)^{\mu_j x}\rho_j{}^x, \cdots,$$

$$\left(\frac{x}{e}\right)^{\mu_n x}\rho_n{}^x e^{2\pi\lambda_{nj}(-1)^{1/2}x}.$$

Dividing by $(x/e)^{\mu_j x}\rho_j{}^x$ and expressing the factors all in the form of exponentials to the base e, we find the exponents to be as follows:

$$2\pi(-1)^{1/2}\left(\lambda_{1j} - \frac{\mu_1 - \mu_j + \log\rho_j - \log\rho_1 - (\mu_1 - \mu_j)\log x}{2\pi(-1)^{1/2}}\right)x, \cdots, 0, \cdots,$$

$$2\pi(-1)^{1/2}\left(\lambda_{nj} - \frac{\mu_n - \mu_j + \log\rho_j - \log\rho_n - (\mu_n - \mu_j)\log x}{2\pi(-1)^{1/2}}\right)x. \tag{21}$$

We discuss separately the cases in which j is associated with the segment of L farthest to the left and the cases in which it is not. Let us begin with the first of these cases.

For values of i not associated with this first segment of L, μ_i is $< \mu_j$ and the exponent

$$(22) \qquad 2\pi(-1)^{1/2}\left(\lambda_{ij} - \frac{\mu_i - \mu_j + \log \rho_j - \log \rho_i - (\mu_i - \mu_j)\log x}{2\pi(-1)^{1/2}}\right)x$$

has a real part which becomes negatively infinite as x recedes to ∞ in any manner whatever in the sector $0 \leqq \arg x \leqq \pi/2$. The terms of (20) for these values of i can therefore play no part in the asymptotic form of $g_{ij}(x)$ in the region under consideration. For values of i associated with this segment of L, $\mu_i = \mu_j$ and the coefficient of x in the ith exponent of the set (21) is a constant. Let the points $P_i (i = 1, 2, \cdots, m)$ represent these coefficients in the complex plane; by virtue of our definition of λ_{ij} all these points except P_j, which is at the origin, lie above the axis of reals and below or upon the line $v = 2\pi$. Furthermore, by the ordering of the ρ_i the points of P_i proceed from right to left as i increases. To allow x to become infinite along a ray $\arg x = \tau$ is precisely the same thing, so far as the behavior of the first m exponents in (21) is concerned, as to rotate the whole set of points P_i bodily through the angle τ and then allow x to become infinite along the positive axis of reals; that point of the set which in the rotated position lies farthest to the right corresponds to the dominant term. Let us construct, then, the broken line $P_1, \cdots, P_j, \cdots, P_m$ above or upon which all the points P_i lie and the vertices of which are points of the set. We are interested in what happens as x becomes infinite in the first quadrant; i.e., in values of τ between 0 and $\pi/2$. The dominant terms correspond, therefore, to the points P_i which are vertices of the broken line $P_1, \cdots, P_m$. Let the acute angles which the successive segments of the portion $P_1, \cdots, P_j$ of this broken line, as we proceed from right to left, make with the axis of reals be denoted by $\phi_1, \phi_2, \cdots$. Then the rays

$$\frac{\pi}{2} - \phi_1, \quad \frac{\pi}{2} - \phi_2, \cdots$$

are, in increasing angular order, the critical rays along each of which the asymptotic form of $g_{ij}(x)$ changes *uniformly in the sense of Birkhoff*; i.e., along the ray $\arg x = \pi/2 - \phi_a$, the form changes uniformly from

$$s_{ir}(x)e^{2\pi\lambda_{rj}(-1)^{1/2}x}c_{rj} \quad \text{to} \quad s_{it}(x)e^{2\pi\lambda_{tj}(-1)^{1/2}x}c_{tj}.$$

Thus the behavior of the solutions associated with the segment of L farthest

to the left is wholly analogous to the behavior of all the solutions in the regular case.

We turn our attention now to the case in which j is not associated with the segment of L farthest to the left. From (22) it is clear that as x becomes infinite in any manner whatsoever in the sector $0 \leqq \arg x \leqq \pi/2 - \epsilon$, the terms corresponding to values of i that belong to the first segment of L on the left dominate all other terms. If there is only one solution associated with that segment of L, it follows therefore that we have

$$g_{ij}(x) \sim s_{i1}(x) e^{2\pi\lambda_{1j}(-1)^{1/2}x} c_{1j} \qquad (j = 2,3, \cdots, n)$$

in the sector $\epsilon \leqq \arg x \leqq \pi/2 - \epsilon$. If there are several solutions, corresponding to $i = 1, 2, \cdots, m$, associated with that segment, we must determine which one dominates. The real part of (22) may be written (setting $x = u + iv$; u, v real):

$$(23) \quad f_i(u,v) = u \log |\rho_i| + u[(\mu_i - \mu_j)(\log |x| - 1) - \log |\rho_j|] - 2\pi\lambda_{ij} v - v \arg \rho_i + v[\arg \rho_j - (\mu_i - \mu_j) \arg x].$$

The quantities in brackets are the same for $i = 1, 2, \cdots, m$; hence along any particular ray $\arg x = \tau$, or $u = sv$, within the first quadrant that term will dominate for which

$$s \log |\rho_i| - 2\pi\lambda_{ij} - \arg \rho_i$$

is largest. For simplicity let us set

$$\lambda_{ij} - \alpha_{ij} = a_{ij},$$

and

$$b_{kl} = \frac{2\pi(\lambda_{kj} - \lambda_{lj}) + \arg(\rho_k/\rho_l)}{\log |\rho_l/\rho_k|} = \frac{2\pi(a_{kj} - a_{lj})}{\log |\rho_l/\rho_k|};$$

we observe $b_{kl} = b_{lk}$. Then it follows from the ordering of the ρ_i according to descending absolute value that the first term of (20) dominates when s satisfies all the inequalities

$$s > b_{1k} \qquad (k = 2,3, \cdots, m).$$

Secondly, if s satisfies the conditions

$$s < b_{12} \text{ and } s > b_{2k} \qquad (k = 3,4, \cdots, m),$$

the second term of (20) dominates. Thus we see that if any of the quantities b_{2k} are $\geqq b_{12}$, there is no region in which the second term dominates. Again, if we have

$$s < b_{13}, \; s < b_{23}, \text{ and } s > b_{3k} \qquad (k = 4,5, \cdots, m),$$

the third term dominates; and so on. Finally, if we have

$$s < b_{mk} \qquad (k = 1,2, \cdots, m-1),$$

the mth term dominates. It is clear that if any of the b_{mk} are $\leqq 0$, the mth term cannot dominate along any ray in the first quadrant. The rays determined by setting s successively equal to

$$b_{12},\ b_{13},\ b_{23},\ \text{etc.}$$

constitute a set of rays of which some may be critical; along one of these critical rays, if there are any, the asymptotic form of $g_{ij}(x)$ changes uniformly from one of the first m terms of (20) to another. Any one of these rays may not be critical in a given instance; or it may happen in a given case that the b_{1k} $(k=2,3, \cdots, m)$ are all negative or zero, so that the first term dominates along any ray in the first quadrant (exclusive of its boundaries) and there are no critical rays in the quadrant. It is clear that *in any event the number of critical rays is* $\leqq m-1$.

The asymptotic form of $g_{ij}(x)$ along any ray R through the origin within the first quadrant and above the uppermost critical ray, however close R may be to the positive v-axis but different from it, is given by one of the first m terms of (20). Yet along that axis itself we have $g_{ij}(x) \sim s_{ij}(x)$. How does the form change from $s_{ij}(x)$ to the other term of (20)? We now investigate this question.

As x becomes infinite along any ray through the origin inclined only slightly to the right of the positive v-axis, the quantities (22) for which $\mu_i > \mu_j$ have real parts all of which are positive for x sufficiently far out on this ray, whereas on the axis itself these real parts are all negative, by our choice of λ_{ij}. This indicates that for each value of i for which $\mu_i > \mu_j$ there is a curve C_i which as it recedes from the origin ultimately lies in any preassigned sector $\pi/2 - \epsilon \leqq \arg x \leqq \pi/2$, however small ϵ may be, and to the left of which (23) is negative and to the right of which (23) is positive. The equation of C_i is of course $f_i(u,v) = 0$. The fact that along any ray in the first quadrant not parallel to the v-axis $f_i(u,v)$ is ultimately positive for the values of i in question, while along any ray in that quadrant parallel to the v-axis $f_i(u, v)$ is ultimately negative, shows that the curve C_i has no rectilinear asymptote.

To examine further the character of C_i let us seek its intersection with the line $u = sv (s > 0)$. If we solve for $\log |x|$ and make use of the relation

$$2\pi\lambda_{ij} + \arg \rho_i - \arg \rho_j = 2\pi a_{ij} - (\mu_i - \mu_j)\pi/2,$$

we find

$$\log|x| = \log\left|\frac{\rho_j}{\rho_i}\right|^{1/(\mu_i-\mu_j)} + \frac{2\pi a_{ij}}{s(\mu_i-\mu_j)} + 1 - \frac{\pi/2 - \arg x}{s}. \tag{24}$$

Inasmuch as s is the tangent of $\pi/2-\arg x$, we see that as s approaches zero and $\arg x$ approaches $\pi/2$, $\log|x|$ and therefore $|x|$ increases indefinitely. Moreover $(\pi/2-\arg x)/s$ approaches 1, so that $\log|x|$ is asymptotic to the sum of the first two terms on the right in (24). In other words, the curve C_i is asymptotic to the following curve in which (r, θ) are the polar coördinates of x, the positive v-axis being taken as the polar axis with the pole at the origin:

$$r = \left|\frac{\rho_j}{\rho_i}\right|^{1/(\mu_i-\mu_j)} e^{-2\pi a_{ij}/((\mu_i-\mu_j)\tan\theta)}. \tag{25}$$

Differentiating (23) we find the slope of C_i to be

$$\frac{dv}{du} = \frac{(\mu_i-\mu_j)\log(u^2+v^2)^{1/2} + \log|\rho_i/\rho_j|}{2\pi a_{ij} + (\mu_i-\mu_j)(\tan^{-1}(v/u) - \pi/2)},$$

which is positive for values of $\tan^{-1}(v/u)$ only slightly less than $\pi/2$; hence the curve C_i, after it has receded from the origin a sufficient distance, rises monotonically toward the right.

Of the curvilinear asymptotes (25) that which, in a region sufficiently remote from the origin, lies farthest to the left is given by the value of i for which

$$\frac{2\pi a_{ij}}{\mu_i-\mu_j} \tag{26}$$

is least; we assume for the present that this quotient takes on its minimum for only one value of i and call this value k. The order of the curves C_i from left to right is according to increasing values of this quotient.

We propose to examine the behavior of $f_i(u,v)$ as x becomes infinite along certain curves of the type of (25); i.e., along the curves $r=e^{-l/\theta}(l>0)$. Changing the variables (u, v) to the polar coördinates (r, θ), we find

$$f_i(u,v) = \phi_i(r,\theta) = r(\mu_j-\mu_i)\sin\theta\left\{\log\left[r \div \left|\frac{\rho_j}{\rho_i}\right|^{1/(\mu_i-\mu_j)} e^{-2\pi a_{ij}/((\mu_i-\mu_j)\tan\theta)}\right] + \frac{\theta}{\tan\theta} - 1\right\}.$$

First let l be any positive constant less than the smallest of the quotients (26). Then the curve $r=e^{-l/\theta}$ lies to the left of all the curves C_i and as x

becomes infinite along this curve, or as θ approaches zero through negative values, we have

$$\phi_i(r,\theta) = \phi_i(e^{-l/\theta},\theta) \to e^{-l/\theta}(\mu_i - \mu_j)\left(l - \frac{2\pi a_{ij}}{\mu_i - \mu_j}\right).$$

Thus we see that along any such curve the real part of the exponent (22), for values of i in which we are interested, becomes negatively infinite, and the asymptotic form of $g_{ij}(x)$ is $s_{ij}(x)$.

If l be taken as any constant greater than the least of the quantities (26) but less than the next larger one, it is clear that along the curve $r = e^{-l/\theta}$, the real part of the exponent (22) given by the value $i = k$ becomes positively infinite, whereas all the others become negatively infinite in real part. Thus along such a curve the asymptotic form of $g_{ij}(x)$ is given by the kth term of (20). If l be taken larger than the next to the smallest quotient (26), two or more of the exponents (22) have real parts that become positively infinite along this curve and we have to determine their relative magnitudes. To this end consider the difference $f_k(u, v) - f_i(u, v)$; for values of i for which $\mu_k = \mu_i$, this reduces to

$$u \log \left|\frac{\rho_k}{\rho_i}\right| - 2\pi v(a_{kj} - a_{ij}). \tag{27}$$

Along any curve for which v/u becomes infinite, such as $r = e^{-l/\theta}$, the second term in (27) dominates. But since $a_{kj} - a_{ij} < 0$ when $\mu_k = \mu_i$, this term is positive and $f_k(u, v)$ dominates $f_i(u, v)$. Secondly for values of i for which $\mu_k \neq \mu_i$, we change the variables to (r, θ) and find

$$f_k(u,v) - f_i(u,v) = \phi_k(r,\theta) - \phi_i(r,\theta)$$

$$= r(\mu_i - \mu_k)\sin\theta\left\{\log\left[r \div \left|\frac{\rho_k}{\rho_i}\right|^{1/(\mu_i-\mu_k)} e^{2\pi(a_{kj}-a_{ij})/((\mu_i-\mu_k)\tan\theta)}\right] + \frac{\theta}{\tan\theta} - 1\right\}.$$

As x becomes infinite along the curve $r = e^{-l/\theta}$, we have

$$f_k(u,v) - f_i(u,v) \to e^{-l/\theta}(\mu_k - \mu_i)\left(l - \frac{2\pi(a_{kj} - a_{ij})}{\mu_k - \mu_i}\right). \tag{28}$$

When $\mu_i < \mu_k$, $f_k(u, v)$ dominates $f_i(u, v)$ for l greater than

$$\frac{2\pi a_{kj}}{\mu_k - \mu_j}, \tag{29}$$

since this ratio is now greater than

$$\frac{2\pi(a_{kj} - a_{ij})}{\mu_k - \mu_i}. \tag{30}$$

When $\mu_i > \mu_k$, (29) is less than (30), and for values of l between the two, $f_k(u, v)$ dominates $f_i(u, v)$.

The above analysis makes it clear that along the curve C_i farthest to the left the asymptotic form of $g_{ij}(x)$ shifts from $s_{ij}(x)$ to the kth term of (20). When μ_k is the slope of the segment of L farthest to the left, the new asymptotic form is valid out to the uppermost critical *ray*, for the asymptotic form cannot shift to a term of (20) associated with another segment of L than the one farthest to the left and it can shift to another term of (20) associated with the same segment of L only along a certain ray, whose equation is obtained by setting (27) equal to zero. If μ_k is not the slope of the first segment of L on the left, the new form holds to the right as far as the curve

$$r = e^{-2\pi(a_{kj} - a_{pj})/((\mu_k - \mu_p)\theta)}, \tag{31}$$

where p is the value of i for which (30) takes on its least value exceeding (29). Along this curve the form changes to the pth term of (20). If p is associated with the segment of L farthest to the left, the present asymptotic form holds clear out to the uppermost critical ray; otherwise, it is valid to the right as far as another critical curve (of the same type as (31)) where a further change takes place, and so on. Each time the form changes, as we go toward the right, the new asymptotic form is a term of (20) associated with a segment of L to the left of that to which the former dominant term belonged. We conclude, therefore, that *the number of critical curves is at least* 1 *and is less than or equal to the number of segments of L lying to the left of that with which j is associated.*

In the event of the quotient (26) taking on its minimum value for two values of i, say k and k', we may ascertain which term dominates by inspection of (27) or (28). If $\mu_k = \mu_{k'}$, (27) shows that the term for which $|\rho_i|$ is greater is the dominating one. It should be observed that $|\rho_k|$ and $|\rho_{k'}|$ cannot be equal, for in the case in question a_{kj} and $a_{k'j}$ are the same; hence by definition of the λ_{ij}, we have $\arg \rho_k = \arg \rho_{k'}$, and if the absolute values of these ρ's were also equal, we should have a multiple root, which is contrary to our hypotheses. If $\mu_k \neq \mu_{k'}$, (28) shows that the dominating term is the one for which μ_i is greater.

Since the definition of asymptotic form is commonly given for a region,* it may be well to make a statement in terms of sectors concerning the asymp-

* Cf Birkhoff, loc. cit., p. 248.

totic form of $g_{ij}(x)$ in the first quadrant. Denoting by arg $x=\alpha$, arg $x=\beta$, $\cdots$, arg $x=\gamma$ the critical rays in increasing angular order, we have

$$g_{ij}(x) \sim s_{ip}(x)e^{2\pi\lambda_{pj}(-1)^{1/2}x} \quad \begin{cases} p = r \text{ for } \epsilon \leqq \arg x \leqq \alpha - \epsilon ; \\ p = t \text{ for } \alpha + \epsilon \leqq \arg x \leqq \beta - \epsilon ; \\ p = w \text{ for } \gamma + \epsilon \leqq \arg x \leqq \pi/2 - \epsilon ; \end{cases}$$

where ϵ is arbitrarily small and positive.

A detailed study of the asymptotic form of $g_{ij}(x)$ in the sector $3\pi/2-\epsilon \leqq \arg x < 2\pi$ lacks interest because in the last section it was not feasible to determine the ν_{ij} in terms of the μ's and ρ's and because there are no fundamental differences between the situations in this region and in the first quadrant. We merely remark that there are, in general, both critical rays and critical curves along which the form of $g_{ij}(x)$ changes from one to another of the terms of the sum

$$s_{ij}(x)e^{2\pi(\nu_{ij}-\mu_1)(-1)^{1/2}x}e^{-2\pi r_1(-1)^{1/2}}d_{1j} + \cdots + s_{in}(x)^{2\pi(\nu_{nj}-\mu_n)(-1)^{1/2}x}e^{-2\pi r_n(-1)^{1/2}}d_{nj},$$

the $s_{ij}(x)$ $(j=1, 2, \cdots, n)$ standing for the same determination of the series here as in (20).

In the cases of Class 2b, §2, the series associated with segments of L having fractional slopes *may* contain no exponential factors $e^{L^{(p)}(x)}$. If such is the case, the work of this section (as well as that of §§2 and 3) is unchanged save for the fact that it must be borne in mind that some or all of the μ's are fractional. If exponential factors do occur, but if none are present in the series corresponding to the segment of L farthest to the left, the situation as regards critical *rays* is no different from that described above. If exponential factors occur in the series associated with the segment of L farthest to the left, then each critical ray is in general replaced by a critical curve which, as it recedes from the origin, ultimately lies in any given small sector enclosing the ray which, were it not for the presence of the exponential factors, would be critical; *these critical curves are, however, algebraic rather than transcendental.* If exponential factors occur in the series associated with the segment of L to which j belongs or with a segment to the left of it, the position and nature of these critical curves lying in the sector $\pi/2-\epsilon \leqq \arg x \leqq \pi/2$ are in general altered; the investigation of the precise nature of the alteration involves considerable algebraic difficulties and we shall not pause to go into it.

CASES IN WHICH SOME OR ALL OF THE CHARACTERISTIC EQUATIONS HAVE MULTIPLE ROOTS

Of the cases in this class that in which the equation (3) has n equal roots, finite and different from zero and not satisfying the secondary equation, is to a large extent typical. The next three sections are devoted to the development of existence theorems for this case; the discussion is followed by a section describing the respects in which this case exemplifies the cases in which roots of different multiplicities occur and pointing out the nature of the existence theorems which can be proved in those cases.

5. **The case of an n-fold multiple root; formal series.** We denote the n-fold root by ρ and assume ρ not a root of (5); the equation (1) is then satisfied formally by the following n series:

$$(32)\quad s_i(x) = \rho^x e^{\gamma_i^{(n-1)}x^{(n-1)/n}+\gamma_i^{(n-2)}x^{(n-2)/n}+\cdots+\gamma_i' x^{1/n}} x^r (1 + s_i' x^{-1/n} + s_i'' x^{-2/n} + \cdots) \qquad (i = 1,2,\cdots,n).$$

In the calculation of the constants use should be made of the expansion

$$\begin{aligned} e^{\gamma^{(n-k)}(x+j)^{(n-k)/n}-\gamma^{(n-k)}x^{(n-k)/n}} &= e^{\gamma^{(n-k)}x^{(n-k)/n}[(1+j/x)^{(n-k)/n}-1]} \\ &= e^{\gamma^{(n-k)}x^{(n-k)/n}[((n-k)/n)(j/x)+\cdots]} \\ &= 1 + \frac{\gamma^{(n-k)}(n-k)j}{nx^{k/n}} + \cdots \\ &\qquad (j = 1,2,\cdots,n;\ k = 1,2,\cdots,n-1) \end{aligned}$$

and of the relations

$$n^{k-1} - (n-1)^k + \frac{(n-1)(n-2)^k}{2!} - \cdots \pm (n-1)2^{k-1} \mp 1 \begin{cases} = 0, & k < n; \\ > 0, & k \geqq n; \end{cases}$$

which are readily established (k being a positive integer). The $\gamma_i^{(n-1)}$ are determined by the equation

$$(33)\quad [\gamma_i^{(n-1)}]^n = -\frac{a_{01}\rho^n + a_{11}\rho^{n-1} + \cdots + a_{n-1,1}\rho + a_{n1}}{\left[n^{n-1} - (n-1)^n + \frac{(n-1)(n-2)^n}{2!} - \cdots \pm (n-1)2^{n-1} \mp 1\right] a_{00}\rho^n} - \frac{n!\, n^{n-1}}{(n-1)^n}.$$

As soon as we fix upon a determination for $\gamma_i^{(n-1)}$, the values of the subsequent constants in the ith series (32) are uniquely determined.

For simplicity we shall throughout our work take that determination of $\arg \gamma_i^{(n-1)}$ which is positive or zero and less than 2π; if another determination were selected, the reasoning would proceed in the same manner. $\gamma_i^{(n-1)}$

will then be chosen, for convenience, to denote that one of the $\gamma_i^{(n-1)}$ whose argument is least. *We assign subscripts to the other $\gamma_i^{(n-1)}$ so that we have**

$$\begin{aligned} &\gamma_2^{(-1)} = \omega_1\gamma_1^{(-1)}, \qquad \gamma_3^{(-1)} = \omega_{n-1}\gamma_1^{(n-1)}, \\ &\gamma_4^{(n-1)} = \omega_2\gamma_1^{(-1)}, \qquad \gamma_5^{(-1)} = \omega_{n-2}\gamma_1^{(n-1)}, \text{ etc.}, \end{aligned} \tag{34}$$

where ω_i $(i=1, 2, \cdots, n-1)$ are the complex nth roots of unity in the order in which one meets them in traversing the unit circle in counter-clockwise direction beginning at 1.

Let us set

$$g_i(x) = g(x + i - 1) \qquad (i = 1,2, \cdots, n)$$

and seek to obtain analytic solutions of equation (1) by obtaining them for the system

$$\begin{aligned} &g_i(x + 1) = g_{i+1}(x) \qquad (i = 1,2, \cdots, n - 1), \\ &g_n(x + 1) = -\frac{a_n(x)}{a_0(x)}g_1(x) - \frac{a_{n-1}(x)}{a_0(x)}g_2(x) - \cdots - \frac{a_1(x)}{a_0(x)}g_n(x), \end{aligned} \tag{35}$$

or in matrix notation,

$$G(x + 1) = A(x)G(x). \tag{36}$$

This equation possesses the formal matrix solution

$$\begin{aligned} S(x) = (s_{ij}(x)) &= \begin{vmatrix} s_1(x) & \cdots & s_n(x) \\ s_1(x+1) & \cdots & s_n(x+1) \\ \cdots & \cdots & \cdots \\ s_1(x+n-1) & \cdots & s_n(x+n-1) \end{vmatrix} \\ &= \rho^x x^r \begin{vmatrix} e^{L_1(x)}\left(1+\frac{s_1'}{x^{1/n}}+\cdots\right) & \cdots & e^{L_n(x)}\left(1+\frac{s_n'}{x^{1/n}}+\cdots\right) \\ e^{L_1(x)}(\rho+\cdots) & \cdots & e^{L_n(x)}(\rho+\cdots) \\ \cdots & \cdots & \cdots \\ e^{L_1(x)}(\rho^{n-1}+\cdots) & \cdots & e^{L_n(x)}(\rho^{n-1}+\cdots) \end{vmatrix}, \end{aligned} \tag{37}$$

* It would seem natural to assign subscripts to the other $\gamma_i^{(n-1)}$ so that we should have

$$\gamma_i^{(n-1)} = \omega_{i-1}\gamma_1^{(n-1)} \qquad (i = 2, 3, \cdots, n).$$

If that assignment were made, we should find that upon replacing $x^{1/n}$ by $\omega_1 x^{1/n}$, $s_1(x)$ would become $s_n(x)$ and likewise $s_i(x)$ would become $s_{i-1}(x)$ $(i=2, 3, \cdots, n)$; indeed we should have

$$\gamma_i^{(n-j)} = \omega_{i-j}\gamma_1^{(n-j)} \qquad (j = 2, 3, \cdots, n - 1)$$

and

$$s_i^{(j)} = \omega_{i-j}s_1^{(j)} \qquad (j = 1, 2, \cdots),$$

a set of relations which might be obtained directly by examination of the expressions for the γ's and s's in terms of the a's of the coefficient functions (2). It is thus clear that the series in (32) are really the n determinations of a single series. On the other hand the assignment of subscripts in this way would make it necessary for us to order the formal series in the matrix $S(x)$, presently to be introduced, in other than the natural order of the subscripts employed in the series, and this would produce considerable inconvenience in the work to follow.

where for brevity we have set

$$(38)\quad L_j(x) = \gamma_j^{(n-1)}x^{(n-1)/n} + \gamma_j^{(n-2)}x^{(n-2)/n} + \cdots + \gamma_j' x^{1/n}\ (j=1,2,\cdots,n),$$

there being no possibility of ambiguity in this section and the following one if the superscript (n) is omitted from $L_j(x)$. Our proof of the existence of analytic solutions of (35) will be based upon modifications of the following infinite products, which are symbolic solutions of that equation:

$$(39)\quad \begin{aligned} &A(x-1)A(x-2)A(x-3)\cdots, \\ &A^{-1}(x)A^{-1}(x+1)A^{-1}(x+2)\cdots. \end{aligned}$$

The n columns of the matrix $S(x)$ are linearly independent sets of functions, since the determinant of the matrix is

$$(40)\quad |S(x)| = \rho^{nz}x^{nr-(n-1)/2}(d+d'x^{-1/n}+\cdots),\ d = C[\gamma_1^{(n-1)}]^{n(n-1)/2} \neq 0,$$

C being a constant not zero. In general the series in $S(x)$ do not converge, but it will be shown that in general there exist two solutions of the system (35) which are analytic over the entire finite x-plane and which are asymptotically represented in a portion of the plane by certain elements of the matrix $S(x)$.

6. **The case of an n-fold multiple root; determinant limits.** We have observed that $S(x)$ is a formal matrix solution of (36),

$$S(x+1) = A(x)S(x).$$

The element in the ith row and jth column of the inverse matrix $S^{-1}(x)$, being the quotient of the cofactor of the element in the jth row and ith column of $S(x)$ by the determinant $|S(x)|$ (cf. (40)), is

$$(41)\quad \bar{s}_{ij}(x) = \rho^{-z}e^{-L_i(x)}x^{(n-1)/2-r}(\bar{s}_{ij} + \bar{s}_{ij}'x^{-1/n} + \bar{s}_{ij}''x^{-2/n} + \cdots).$$

Let $T(x)$ denote the matrix obtained from $S(x)$ by replacing $s_{ij}(x)$ by $t_{ij}(x)$, where the series in the latter is convergent for x in the neighborhood of ∞ and has the same terms as the series in $s_{ij}(x)$ up to and including that in $x^{-(K-1)/n}$. We may then define a matrix $B(x)$ by the relation

$$T(x+1) = B(x)T(x).$$

$B(x)$ is a matrix of functions $b_{ij}(x)$ each of which has an expansion in powers of $x^{-1/n}$ that agrees precisely with $a_{ij}(x)$ (the element in the ith row and jth column of $A(x)$) up to and including the term in $x^{-(K-1)/n}$. This follows upon comparing

$$B(x) = T(x+1)T^{-1}(x) \quad \text{and} \quad A(x) = S(x+1)S^{-1}(x)\ ;$$

for $T(x+1)$ is the same as $S(x+1)$ up to and including the term in $x^{-(K-1)/n}$ and similarly for $T^{-1}(x)$ and $S^{-1}(x)$, because $\bar{l}_{ij}(x)$ of $T^{-1}(x)$ is given by (41) up to and including terms of degree $K-1$ in $x^{-1/n}$. We therefore have

$$A(x) = B(x) + M(x),$$

where $M(x)$ is a matrix of power series in $x^{-1/n}$ whose lowest degree terms are of degree K or higher and which converge in the vicinity of ∞. Then $N(x)$, defined by

$$M(x) = B(x)N(x),$$

is a matrix of power series in $x^{-1/n}$ of the same type as those in $M(x)$. Hence we have

$$A(x) = B(x)[I + x^{-K/n}C(x)]$$

in which I is the unit matrix and $C(x)$ is a matrix of power series in $x^{-1/n}$ convergent for $|x|>R$.

We shall understand the definition* of asymptotic representation of a function with respect to x or v $(x=u+(-1)^{1/2}v)$ by a power series in x^{-1} to be modified in the natural manner to define representation by a series of the type of those in (32).

For convenience in stating the following theorem let us denote by ϕ the principal determination of

$$\arg\,[\gamma_1{}^{(-1)}]^n\,;$$

for the present we assume

$$0 < \phi < \pi.$$

THEOREM A. *Form the product of matrices*

$$P_m(x) = A(x-1)A(x-2)\cdots A(x-m)T(x-m).$$

If a suitable determination of $x^{1/n}$ be selected and m be allowed to become infinite, each λ-rowed determinant formed from the first λ columns $(\lambda=1, 2, \cdots, n)$ of $P_m(x)$ converges, for K sufficiently large, to a definite limit function $u_{ij\cdots l}(x)$, independent of K. These limit functions are analytic over the entire finite plane except at points which are singularities of $A(x-1)$ or are congruent on the right† to them. For odd [even] values of λ less than n the limit functions

* Cf. Birkhoff, loc. cit., p. 248.

† The points $x+1, x+2, x+3, \cdots$ [$x-1, x-2, x-3, \cdots$] are said to be congruent on the right [left] to the point x.

are asymptotically represented with respect to x in the sector $\pi/2 \leqq \arg x \leqq \pi + \phi - \epsilon$ $[\phi + \epsilon \leqq \arg x \leqq 3\pi/2]$ (ϵ being an arbitrarily small positive quantity) by the corresponding determinant $s_{ij\ldots l}(x)$ of $S(x)$. The asymptotic form of $u_{ij\ldots l}(x)$ for λ odd [even] and less than n is given by $s_{ij\ldots l}(x)$ with respect to v in the sector $0 < \arg x < \pi/2$ $[3\pi/2 < \arg x < 2\pi]$. The determinant $|P_m(x)|$ (given by $\lambda = n$) is asymptotically represented by the determinant $|S(x)|$ with respect to x in the sector $\pi/2 \leqq \arg x \leqq 3\pi/2$ and with respect to v in the sectors $0 < \arg x < \pi$ and $3\pi/2 < \arg x < 2\pi$.

We may write

$$P_m(x) = T(x)\overline{P}_m(x),$$

where

$$\overline{P}_m(x) = [T^{-1}(x)A(x-1)T(x-1)][T^{-1}(x-1)A(x-2)T(x-2)] \cdots [T^{-1}(x-m+1)A(x-m)T(x-m)].$$

The elements of $T(x)$ are polynomials or series convergent for $|x| > R$. In order to show that the determinants in $P_m(x)$ converge to definite limit functions as m becomes infinite, it will be sufficient to prove that the corresponding determinants in $\overline{P}_m(x)$ so converge. We proceed to show this first for $\lambda = 1$.

The matrix $\overline{P}_m(x)$ is the product of matrices of the type

$$T^{-1}(x+1)A(x)T(x) = T^{-1}(x+1)B(x)T(x) + T^{-1}(x+1)B(x)x^{-K/n}C(x)T(x) = I + x^{-K/n}T^{-1}(x)C(x)T(x),$$

I representing again the unit matrix. The second term is a matrix

$$(x^{-K/n+(n-1)/2}e^{L_j(x)-L_i(x)}\lambda_{ij}(x)) = (\theta_{ij}(x)) = \Theta(x),$$

in which the $\lambda_{ij}(x)$ are power series in $x^{-1/n}$ convergent in the vicinity of $x = \infty$. We shall now suppose K to be taken $\geqq 2n + n(n-1)/2$, so that we have

$$\frac{K}{n} - \frac{n-1}{2} = d \geqq 2. \tag{42}$$

The matrix $\overline{P}_m(x)$ may now be written

$$[I + \Theta(x-1)][I + \Theta(x-2)] \cdots [I + \Theta(x-m)]$$

$$= I + \sum_{k_1=1}^{m} \Theta(x-k_1) + \sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} \Theta(x-k_1)\Theta(x-k_2) + \cdots.$$

The ith element ($i=1, 2, \cdots, n$) in the first column is

$$\begin{aligned}\bar{p}_{m;\,i1}(x) &= \delta_{i1} + \sum_{k_1=1}^{m}\theta_{i1}(x-k_1) \\ &\quad + \sum_{k_1=1}^{m-1}\sum_{k_2=k_1+1}^{m}\Big[\sum_{\tau=1}^{n}\theta_{i\tau}(x-k_1)\theta_{\tau 1}(x-k_2)\Big] + \cdots \\ &= \delta_{i1} + \sum_{k_1=1}^{m}\frac{\lambda_{i1}(x-k_1)}{(x-k_1)^d}\,e^{L_1(x-k_1)-L_i(x-k_1)} \\ &\quad + \sum_{k_1=1}^{m-1}\sum_{k_2=k_1+1}^{m}\Big[\sum_{\tau=1}^{n}\frac{\lambda_{i\tau}(x-k_1)\lambda_{\tau 1}(x-k_2)}{(x-k_1)^d(x-k_2)^d}\,e^{L_\tau(x-k_1)-L_i(x-k_1)+L_1(x-k_2)-L_\tau(x-k_2)}\Big] + \cdots.\end{aligned} \tag{43}$$

But the $\lambda_{ij}(x)$ $(i, j=1, 2, \cdots, n)$ are bounded in the neighborhood of ∞; i.e., $|\lambda_{ij}(x)|$ $(i, j=1, 2, \cdots, n)$ is less than some constant M for $|x|>R$. Therefore, if x lies in the region D of the accompanying figure, so that the points $x-i$ $(i=1, 2, \cdots)$ are all exterior to the circle of radius R about the origin, the product of the λ-functions in the typical element in the $(l+1)$th term of (43),

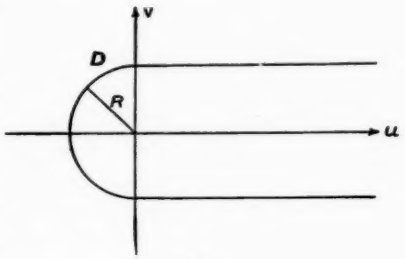

FIG. 2

$$\frac{\lambda_{i\tau}(x-k_1)\lambda_{\tau\sigma}(x-k_2)\cdots\lambda_{\nu 1}(x-k_l)}{(x-k_1)^d(x-k_2)^d\cdots(x-k_l)^d} \cdot e^{L_\tau(x-k_1)-L_i(x-k_1)+L_\sigma(x-k_2)-L_\tau(x-k_2)+\cdots+L_1(x-k_l)-L_\nu(x-k_l)} \tag{44}$$

is less in absolute value than M^l. As for the exponential term, that may be written

$$\begin{aligned}&e^{L_1(x-k_1)-L_i(x-k_1)}e^{L_\tau(x-k_1)-L_\tau(x-k_2)-L_1(x-k_1)+L_1(x-k_2)} \\ &\quad \cdot e^{L_\sigma(x-k_2)-L_\sigma(x-k_3)-L_1(x-k_2)+L_1(x-k_3)} \\ &\quad \cdots e^{L_\nu(x-k_{l-1})-L_\nu(x-k_l)-L_1(x-k_{l-1})+L_1(x-k_l)}.\end{aligned} \tag{45}$$

We propose to show that, given any value of $[\gamma_1{}^{(n-1)}]^n$ above the axis of reals, a suitable choice of the determination of $x^{1/n}$ will make the absolute value of each factor after the first in this product less than or equal to 1 for x in a certain part of the region D.

When τ, or σ, or $\cdots$ is 1, it is clear that the factor of (45) in which it

occurs has the value unity. We need consider, therefore, the factors in which the exponent is

$$(46)\quad [L_j(x-k_r)-L_j(x-k_s)]-[L_1(x-k_r)-L_1(x-k_s)]\begin{pmatrix} j=2,3,\cdots,n \\ r<s \end{pmatrix}.$$

Taking into account the relations (34) and (38), we see that this difference may be written in the form

$$\begin{aligned}(47)\quad &\gamma_1^{(n-1)}(\omega_\alpha-1)\left[(x-k_r)^{(n-1)/n}-(x-k_s)^{(n-1)/n}\right]\\ &+(\gamma_j^{(n-2)}-\gamma_1^{(n-2)})\left[(x-k_r)^{(n-2)/n}-(x-k_s)^{(n-2)/n}\right]\\ &+\cdots+(\gamma_j'-\gamma_1')\left[(x-k_r)^{1/n}-(x-k_s)^{1/n}\right],\end{aligned}$$

$$\alpha=\begin{cases}\dfrac{j}{2} \text{ for } j \text{ even};\\ n-\dfrac{j-1}{2} \text{ for } j \text{ odd}.\end{cases}$$

If R (cf. Figure 2) is sufficiently large, the first term of this sum is in absolute value the dominating one, whatever determination of $x^{1/n}$ be selected, for all x in D and for all possible values of k_r and k_s. We now prove this for the case in which x lies in the upper half-plane (including the part of the negative axis of reals in D) and arg $x^{1/n}$ has its smallest positive value; for x in the lower half-plane and for other determinations of $x^{1/n}$, the situation is wholly similar. The proof is accomplished by showing that, given any constants $C_p(\neq 0)$ and C_{p-1} and an arbitrarily large positive number N, we can take R so great that the inequality

$$\left|\frac{C_p\left[(x-k_r)^{(n-p)/n}-(x-k_s)^{(n-p)/n}\right]}{C_{p-1}\left[(x-k_r)^{(n-p-1)/n}-(x-k_s)^{(n-p-1)/n}\right]}\right|>N \qquad (p=1,2,\cdots,n-2)$$

is satisfied. The left-hand member may be written

$$(48)\quad \frac{C_p}{C_{p-1}}\cdot\frac{(x-k_r)^{(n-p-1)/n}+(x-k_r)^{(n-p-2)/n}(x-k_s)^{1/n}+\cdots+(x-k_s)^{(n-p-1)/n}}{(x-k_r)^{(n-p-2)/n}+(x-k_r)^{(n-p-3)/n}(x-k_s)^{1/n}+\cdots+(x-k_s)^{(n-p-2)/n}}.$$

The numerator of the second quotient is to be examined next. The argument of each term in this sum, for any p, is positive and $\leqq(n-2)\pi/n$. Hence the argument of the whole sum exclusive of the last term is confined to these same limits, and it follows that the absolute value of the whole sum is*

* If $n\leqq 4$, the factor sin $(\pi/2-2\pi/n)$ occurring here is replaced by 1, since each term in the sum, for any p, has an argument between 0 and $\pi/2$ (inclusive of the latter).

$\geqq \sin(\pi/2-2\pi/n)$ times the absolute value of that sum with the last term deleted. Thus the absolute value of the expression (48) is

$$\geqq \left|\frac{C_p}{C_{p-1}}(x - k_r)^{1/n}\right| \sin\left(\frac{\pi}{2} - 2\frac{\pi}{n}\right).$$

But $|(x-k_r)^{1/n}| > R^{1/n}$, which we can make as large as we please by taking R sufficiently large.

The results of the preceding paragraph show that if the first term in the sum (47) is kept away from an arbitrarily small sector enclosing the axis of imaginaries, the real part of this first term will be the dominating influence in determining the real part of the function (46). We want the real part of that function to be negative, in order that e to that power shall be in absolute value less than 1. In other words we desire to have*

$$\begin{aligned}\frac{\pi}{2} + \frac{\epsilon}{n} &\leqq \arg\left\{\gamma_1^{(n-1)}(\omega_\alpha - 1)\left[(x - k_r)^{(n-1)/n} - (x - k_s)^{(n-1)/n}\right]\right\} \\ &\leqq \frac{3\pi}{2} - \frac{\epsilon}{n},\end{aligned} \tag{49}$$

where ϵ is an arbitrarily small positive number.

By definition of ϕ we have

$$0 < \arg \gamma_1^{(n-1)} = \frac{\phi}{n} < \frac{\pi}{n}.$$

It is a simple matter to show that $\arg(\omega_\alpha - 1)$ is bounded as follows:

$$\frac{\pi}{2} + \frac{\pi}{n} \leqq \arg(\omega_\alpha - 1) \leqq \frac{3\pi}{2} - \frac{\pi}{n} \qquad (\alpha = 1, 2, \cdots, n-1).$$

Therefore, if condition (49) is to be satisfied, we must have

$$-\frac{\pi}{n} - \frac{\phi}{n} + \frac{\epsilon}{n} \leqq \arg\left[(x - k_r)^{(n-1)/n} - (x - k_s)^{(n-1)/n}\right] \leqq \frac{\pi}{n} - \frac{\phi}{n} - \frac{\epsilon}{n}. \tag{50}$$

For any particular x the points $x-i$ $(i=1, 2, 3, \cdots)$ stretch away to the left on the parallel through x to the axis of reals. Such a parallel corresponds to a curve of hyperbolic type (an actual hyperbola if $n=2$) in the $x^{1/n}$-plane, as indicated in Figure 3. C_1 is the correspondent (first determination) of a parallel in the upper half of the x-plane; C_2 of one in the lower half. Using a second plane for $x^{(n-1)/n}$, C_1' and C_2' (Figure 4) are the transforms

* Here and in the next few pages we shall understand the principal value of the argument to be meant unless the contrary is specified.

of C_1 and C_2. The vectors are drawn to indicate the value of the arguments of the differences $(x-k_r)^{(n-1)/n}-(x-k_s)^{(n-1)/n}$. From the fact that the mapping is conformal it follows readily that if we have

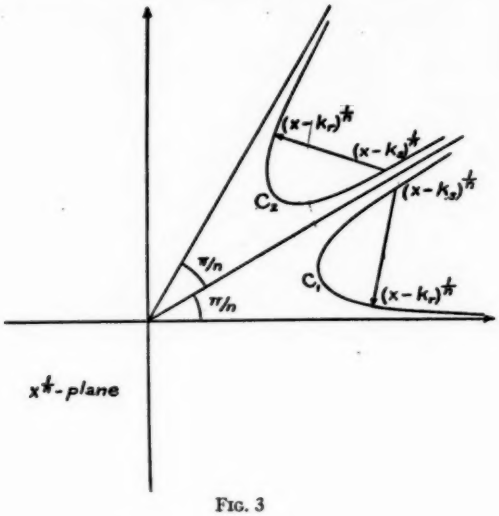

FIG. 3

$$0 \leqq \arg x \leqq \pi + \phi - \epsilon \tag{51}$$

and if we select the first determination* for $x^{1/n}$, the relation (50) will be satisfied. The part of the region D which is also in the sector (51) will hereafter be spoken of as the region D'.

Returning now to the first factor in (45), we see that it is less in absolute value than $e^{L_1(x)-L_i(x)}$, since the ratio

$$\frac{e^{L_1(x-k_1)-L_i(x-k_1)}}{e^{L_1(x)-L_i(x)}} = e^{L_i(x)-L_i(x-k_1)-L_1(x)+L_1(x-k_1)}$$

is of the same type as the factors in (45) after the first, and therefore in absolute value less than unity for all x in D' and for $k_1=1, 2, 3, \cdots$.

We conclude that for x in D' the typical element (44) is less in absolute value than

$$\frac{M^l}{(x-k_1)^d(x-k_2)^d\cdots(x-k_l)^d}\, e^{L_1(x)-L_i(x)},$$

* For brevity we shall speak of the determination of $x^{1/n}$ for which we have $0 \leqq \arg x^{1/n} < 2\pi/n$ as the *first* determination, of the determination for which we have $2\pi/n \leqq \arg x^{1/n} < 4\pi/n$ as the *second* determination, etc.

and hence that the sum (43) is dominated by the sum

$$\delta_{i1} + \left| e^{L_1(x)-L_i(x)} \right| \sum_{k_1=1}^{m} \frac{M}{|x-k_1|^d} + \left| e^{L_1(x)-L_i(x)} \right| \sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} \frac{nM^2}{|x-k_1|^d |x-k_2|^d} + \cdots . \tag{52}$$

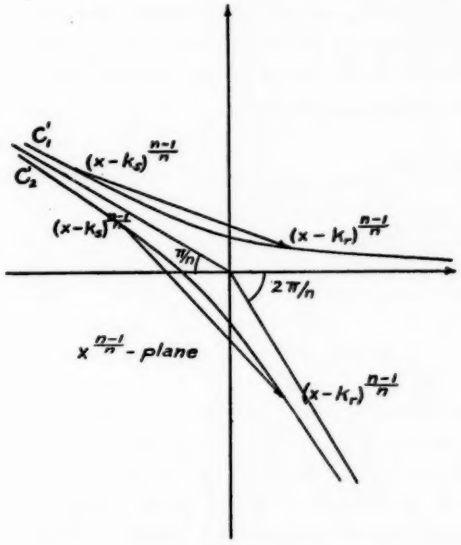

FIG. 4

Allowing m to become infinite, (43) becomes a multiple series and the sum of terms in (52) after the first approaches the limit

$$\frac{1}{n} \left| e^{L_1(x)-L_i(x)} \right| \left\{ \left[\left(1 + \frac{M}{|x-1|^d} \right) \left(1 + \frac{M}{|x-2|^d} \right) \cdots \right] - 1 \right\}.$$

It is thus clear that the elements of the first column of $\overline{P}_m(x)$, and therefore of $P_m(x)$, converge absolutely and uniformly to functions analytic in the vicinity of points in D'. Moreover, it is possible to write

$$P_m(x) = A(x-1)A(x-2) \cdots A(x-r)P_{m-r}(x-r) ;$$

for any finite x not in D', we may take r so large that the points $x-r-1$, $x-r-2$, $\cdots$ will lie in D', and we may carry through for $P_{m-r}(x-r)$ the argument given above for $P_m(x)$. Thus the elements of the first column

of $P_m(x)$ converge absolutely and uniformly in the vicinity of any point of the entire finite plane, excepting only points which are singularities of $A(x-1)$ or are congruent on the right to them, and the limit function $u_i(x)$ to which $p_{m;i1}(x)$ converges is analytic throughout the plane with the exceptions noted.

The proof that the limit functions $u_i(x)$ are independent of K and that they are asymptotically represented by $s_i(x)$ is now carried out as in Birkhoff's paper. The extension of the proof to the cases $\lambda=2, 3, \cdots, n$ also involves little modification of his work. For $\lambda=2$ the region of convergence of the elements of the first column of the new difference system is determined by the fact that we have

$$\frac{\pi}{2}+\frac{2\pi}{n} \leqq \arg(\omega_\alpha+\omega_\beta-1-\omega_1) \leqq \frac{3\pi}{2} \quad (\alpha,\beta=1,2,\cdots,n),$$

in which ω_n is 1, α is $\neq\beta$, and the combination (α, β) is distinct from that of $(n, 1)$. Therefore when λ is 2 the relation (50) is replaced by

$$-\frac{2\pi}{n}-\frac{\phi}{n}+\frac{\epsilon}{n} \leqq \arg\left[(x-k_r)^{(n-1)/n}-(x-k_s)^{(n-1)/n}\right] \leqq -\frac{\phi}{n}-\frac{\epsilon}{n}$$

and the sector (51) by

$$\phi+\epsilon \leqq \arg x \leqq 2\pi.$$

The case of $\lambda=2$ [1] is typical of that in which λ has any even [odd] value $<n$. In the case of $\lambda=n$ no exponential factors (45) occur in the terms of the sum corresponding to (43), since we have

$$\sum_{j=1}^{n} L_j(x) = 0.$$

Hence convergence can be proved as in the regular case and the asymptotic form $|S(x)|$ is valid for the same region as in that case.

The case of $[\gamma_1^{(n-1)}]^n$ *in the lower half-plane.* If we have

$$\arg[\gamma_1^{(n-1)}]^n = \pi+\phi, \qquad 0<\phi<\pi,$$

the only modification that need be made in Theorem A is the interchange, for λ odd and even and $<n$, of the sectors in which the asymptotic forms of the determinant limits are valid. The determination of $x^{1/n}$ which must be selected when λ is odd is again the first; for λ even the second determination must be chosen to secure convergence of the determinants.

The case of $[\gamma_1^{(n-1)}]^n$ *real.* If $[\gamma_1^{(n-1)}]^n$ is negative, the determinant limits exist and are asymptotically represented with respect to x in the sector $\epsilon \leqq \arg x \leqq 2\pi - \epsilon$. If $[\gamma_1^{(n-1)}]^n$ is positive, we cannot be sure of the existence of determinant limits for any determination of $x^{1/n}$.

In an entirely similar manner we may prove a second theorem in which it is necessary to distinguish between two cases according as n is even or odd.

THEOREM B. *Form the product of matrices*

$$P_m'(x) = A^{-1}(x)A^{-1}(x+1)\cdots A^{-1}(x+m-1)T(x+m).$$

If a suitable determination of $x^{1/n}$ *be selected and* m *allowed to become infinite, each* λ*-rowed determinant formed from the last* λ *columns converges, for* K *sufficiently large, to a definite limit function* $v_{ij\cdots l}(x)$*, independent of* K*. These limit functions are analytic over the entire finite plane except at points which are singularities of* $A^{-1}(x)$ *or are congruent on the left to them. For* n *even and* λ *odd* [*even*] *and* $<n$ *the limit functions are asymptotically represented with respect to* x *in the sector* $-\pi+\phi+\epsilon \leqq \arg x \leqq \pi/2$ $[-\pi/2 \leqq \arg x \leqq \phi-\epsilon]$ *by the corresponding determinant* $s_{ij\cdots l}(x)$ *of* $S(x)$*. The asymptotic form of* $v_{ij\cdots l}(x)$ *for* n *even and* λ *odd* [*even*] *and* $<n$ *is given by* $s_{ij\cdots l}(x)$ *with respect to* v *in the sector* $\pi/2 < \arg x < \pi$ $[-\pi < \arg x < -\pi/2]$*. For* λ *odd the determination of* $x^{1/n}$ *chosen is the first for* x *in the upper half-plane, the* n*th for* x *in the lower; for* λ *even the determination chosen is the first in the lower half-plane, the second in the upper. When* n *is odd the sectors for* λ *odd and even are interchanged. The determinant* $|P_m'(x)|$ *is asymptotically represented by the determinant* $|S(x)|$ *with respect to* x *in the sector* $-\pi/2 \leqq \arg x \leqq \pi/2$ *and with respect to* v *in the sectors* $-\pi < \arg x < -\pi/2$ *and* $\pi/2 < \arg x < \pi$.

When $[\gamma_1^{(n-1)}]^n$ is in the lower half-plane or is real, modifications in Theorem B must be made like those indicated above for Theorem A in these cases.

If ρ is a simple root of the secondary equation (5), results analogous to those of Theorems A, B can be obtained. When n is odd the series which contains no exponential factor

$$e^{\gamma_1^{(n-2)}x^{(n-2)/(n-1)}+\cdots}$$

(cf. §1) should be ordered in the $((n+1)/2)$th position, the other series being ordered as above. The only change in the results as stated in Theorem A is the interchange of sectors for $\lambda = (n+1)/2, \cdots, n-1$. When n is even the series from which the exponential factor is missing should be put in the $(n/2+1)$th position. The results are the same except that for $\lambda = n/2$

and $\lambda = n/2+1$ the region of validity of the asymptotic form is somewhat smaller and for $\lambda = n/2+2, \cdots, n-1$ the sectors are interchanged.

No difficulties stand in the way of proving the existence of solutions associated with the determinant limits. Beyond this point, however, the theory cannot be extended by the use of any of the evaluations of the operator Σ, either in terms of series or of contour integrals, that have been used up to the present time in the theory of difference equations. In attempting to obtain intermediate or principal solutions on the left, for example, the exponential factors present make it necessary, in order to secure convergence of the series or of the contour integrals, to proceed *toward the right*, and this cannot be done without going outside the region of validity of the asymptotic form of the functions whose finite integrals we are seeking to determine.

7. **The case of an equation of second order with a double root.** It seems desirable to call particular attention to the case of the equation of second order whose characteristic equation has a double root, for the reason that in this case the above argument leads to somewhat more inclusive results. We observe first of all that, grouping the results of Theorems A and B for $\lambda = 1$, we have one complete set of solutions of the system (35) which are analytic over the entire finite plane and each of which is given asymptotically by the corresponding formal series in a certain portion of the plane.

Secondly it is possible, on the basis of the results of §6 for $n = 2$, to show the existence of intermediate solutions by giving to the operator Σ the series evaluation employed by Birkhoff for the same purpose in the regular case. The elements of the first [second] column of the matrix of intermediate solutions on the left [right], together with their asymptotic properties, are provided by Theorem A[B]; the elements of the second [first] column in general differ from those of the regular case in that they are analytic only above or below some parallel to the real axis—whether above or below depends upon the value of $[\gamma_i']^2$. If this quantity is represented by a point in the upper or lower half-plane, the region of validity of the elements of the columns in question is respectively above or below a suitable parallel to the axis of reals.

Furthermore, there occurs in the work of §6 this essential difference when n is 2: since $L_i(x)$ is a monomial, $\gamma_i' x^{1/2}$, the argument to show the dominance of the leading term of $L_i(x)$ drops out and the ϵ appearing in the inequalities (49), (50), and (51) and in the statement of Theorems A and B is not present. Hence if $[\gamma_i']^2$ takes on a negative real value, the ϕ of Theorem A becomes π and the sector (51) becomes $0 \leqq \arg x \leqq 2\pi$. The elements of the first column therefore have precisely the same properties

as in the regular case except that they are analytic in the *cut* plane, the cut being made from 0 to ∞ along the positive axis of reals; intermediate and principal solutions on the left are then readily obtained. Before allowing $[\gamma_1']^2$ to assume a positive real value let us observe that in the proof of Theorem A there is a sector, $-\pi+\phi+\epsilon \leqq \arg x \leqq 0$, which is, so to speak, unused and apparently unusable; for, although inequality (50) will be satisfied if $x-k_r$ and $x-k_s$ lie in this sector and the nth determination of $x^{1/n}$ is chosen, k_r and k_s do not remain small and as they increase, the points $x-k_r$ and $x-k_s$ are sure to move outside the sector. When n is 2, this sector is $-\pi+\phi \leqq \arg x \leqq 0$. Thus if $[\gamma_1']^2$ is a positive real, so that ϕ is zero, the sector previously used becomes $0 \leqq \arg x \leqq \pi$ and the unused sector becomes $-\pi \leqq \arg x \leqq 0$, both of which are usable. The determinations of $x^{1/2}$ which must be chosen are, however, different for x in the respective half-planes—the first for x in the upper, the second for x in the lower. Hence the limit functions for x in the upper half-plane do not join on analytically along the axis of reals to those for x in the lower half-plane, since the two determinations of $x^{1/2}$ join on continuously along the *positive* axis of reals, which is outside the region D. Intermediate, but not principal, solutions can then be obtained as in the regular case. Parallel statements can be made with reference to solutions on the right when $[\gamma_1']^2$ is real.

8. **Analytic solutions in a large class of irregular cases.** We point out briefly the nature of the results which can be obtained by the methods used above in cases in which the characteristic equation (3) has roots of different multiplicities and a full set of formal series can be found. Our considerations will at first be confined to cases in which the roots of (3) are all finite and non-zero, and for the sake of clarity and simplicity we shall begin with a particular case of rather general type. The facts will be stated only for solutions on the left; the situation with respect to solutions on the right is parallel.

Let the roots ρ_i of (3) be ordered according to decreasing absolute value and let us suppose that

$$\rho_1, \rho_2, \cdots, \rho_k \quad \text{are simple roots;}$$

$$\rho_{k+1} = \rho_{k+2} = \cdots = \rho_l \quad \text{is a multiple root;}$$

$$\rho_{l+1}, \rho_{l+2}, \cdots, \rho_m \quad \text{are simple roots;}$$

$$\rho_{m+1} = \rho_{m+2} = \cdots = \rho_p \quad \text{is a multiple root; and}$$

$$\rho_{p+1}, \rho_{p+2}, \cdots, \rho_n \quad \text{are simple roots.}$$

Let us assume further that we have

$$|\rho_k| > |\rho_{k+1}|\ ,\ |\rho_l| > |\rho_{l+1}|\ ,\ |\rho_m| > ||\rho_{m+1}|\ ,\ \rho_p| > |\rho_{p+1}|.$$

Then the existence of determinant limits like those of the regular case can be proved for $\lambda = 1, 2, \cdots, k, l, l+1, \cdots, m, p, p+1, \cdots, n$; for other values of λ the determinant limits are analytic except for poles over the entire finite plane, but they are asymptotically represented by the corresponding determinants of $S(x)$ only in regions like D' of §6. The extension to solutions associated with the determinant limits is immediate. For $i = 1, 2, \cdots, k$ intermediate and principal solutions may be obtained as in the regular case; for $i = k+1$ an intermediate solution can be gotten like those of §7 if the left-hand member of (5) for $\rho = \rho_{k+1}$ is not real; for $i > k+1$ intermediate solutions can be found only if ρ_{k+1} is a root of multiplicity 2. In this case intermediate solutions as in §7 can be obtained for $i = k+2, \cdots, m$; if the left-hand member of (5) for $\rho = \rho_{k+1}$ is a negative real, principal solutions can be obtained for $i = k+1, \cdots, m$. If the left-hand member of (5) both for $\rho = \rho_{k+1}$ and $\rho = \rho_{m+1}$ has its argument in the open interval $(0, \pi)$ or in the open interval $(\pi, 2\pi)$, an intermediate solution as in §7 can be obtained for $i = m+1$. We can go no further than this unless the multiplicity of the root ρ_{m+1} is 2 and the hypothesis of the preceding sentence is also satisfied. If such is the case, intermediate solutions as in §7 can be found for $i = m+2, \cdots, n$; when the left-hand member of (5) both for $\rho = \rho_{k+1}$ and ρ_{m+1} is a negative real, principal solutions can be found for $i = m+1, \cdots, n$.

It may be remarked further that if a multiple root and a simple root have the same absolute value, the corresponding formal series should be ordered as are the series in the case referred to at the close of §6, in which ρ is a simple root of the secondary equation; determinant limits then exist as in that case. If two multiple roots of different multiplicities have the same absolute value, they should be ordered according to decreasing multiplicity; determinant limits exist as in §6. If two multiple roots have the same absolute value and the same multiplicity, they should be ordered according to decreasing absolute value of the leading γ in the corresponding formal series; i.e., the series corresponding to the root ρ_i for which the left-hand member of (5) has the larger absolute value should be placed first. The points in the complex plane representing each set of γ's are then located at the vertices of a regular polygon whose center is the origin. If the smaller polygon lies within the larger, the determinant limits for λ equal to the value of i which gives the first series of the first set exist as in §6, and if the root yielding the first of the two sets is the first multiple root and is not real, an intermediate

solution can be obtained corresponding to this value of i. Whatever the relative size and position of the two polygons may be, the regions of validity of the asymptotic forms of some or all of the determinant limits for λ greater than this value of i are further restricted.

The extension of these statements to cases of Class 2, §1, in which the characteristic equations associated with one or more of the segments of L have multiple roots is immediate.

BROWN UNIVERSITY,
PROVIDENCE, R. I.

ON THE CONVERGENCE OF QUADRATURE FORMULAS RELATED TO AN INFINITE INTERVAL*

BY

J. V. USPENSKY

The quadrature formulas we shall discuss in this paper belong to the same type as the well known formula given by Gauss. They can be characterized as follows. Let $p(x)$ denote a function which does not assume negative values in a given interval (a, b) and is subject to the condition that the integrals

$$c_m = \int_a^b x^m p(x)dx,$$

or *moments* of the distribution determined by $p(x)$, exist. Under such circumstances it is always possible to find n real numbers

$$x_1 < x_2 < x_3 < \cdots < x_n$$

belonging to the interval (a, b) together with n corresponding constants $A_1, A_2, \cdots, A_n$ such that the formula

$$\int_a^b p(x)f(x)dx = A_1 f(x_1) + A_2 f(x_2) + \cdots + A_n f(x_n) \tag{1}$$

holds true whenever $f(x)$ is a polynomial whose degree does not exceed $2n-1$, and this property completely determines the numbers $x_1, x_2, \cdots, x_n$ as well as the corresponding coefficients $A_1, A_2, \cdots, A_n$. It is important to notice that these coefficients are all *positive*. For any function $f(x)$ which is not a polynomial of degree $\leqq 2n-1$ the right member of (1) ceases to represent exactly the integral in the left member, and we must add the remainder R_n in order to have an exact equality:

$$\int_a^b p(x)f(x)dx = A_1 f(x_1) + A_2 f(x_2) + \cdots + A_n f(x_n) + R_n. \tag{2}$$

For any function $f(x)$ possessing a derivative of order $2n$ the following expression for the remainder R_n can be obtained:

$$R_n = \frac{f^{(2n)}(\xi)}{2n!} \int_a^b p(x)\omega_n^2(x)dx, \tag{3}$$

* Presented to the Society, April 16, 1927; received by the editors in September, 1927.

where

$$\omega_n(x) = (x - x_1)(x - x_2) \cdots (x - x_n),$$

and ξ denotes a certain number belonging to the interval (a, b). For a given particular function and for a given value of n this expression may be useful in that it enables us to form an idea of the accuracy attained by using the approximate formula (1). But it is hardly adapted to yield any conclusion as to the behavior of R_n when n increases indefinitely. Looking at the question from this point of view the problem naturally arises to investigate whether the approximate formula (1) can be made to furnish an indefinite approximation by choosing the number n, which is at our disposal, sufficiently large, or, in other words, whether the remainder R_n in (2) converges to zero with indefinitely increasing n. This important problem of convergence of quadrature formulas of Gaussian type has been almost completely solved in case of the finite interval (a, b) by Stieltjes in his beautiful paper *Quelques recherches sur la théorie des quadratures dites mécaniques*.* The case of an infinite interval is more difficult, and although we possess a very profound investigation concerning the special case

$$f(x) = \frac{1}{z - x}$$

by Stieltjes,† it seems that this problem has never been considered for a more general type of functions $f(x)$ except in an article by the author published in 1916.‡

In the present paper we endeavor to treat the same problem by a totally different method, which permits us to reach more general conclusions than those established in the above mentioned article.

1. We begin with the case in which $b = +\infty$ while a is any finite number, and, to avoid unnecessary complications, we suppose at first $a = 0$. In this case all the numbers $x_1, x_2, \cdots, x_n$ are positive and furthermore the following important inequality holds:

$$L_{m,n} = A_1 x_1^m + A_2 x_2^m + \cdots + A_n x_n^m \leqq c_m \qquad (m = 0, 1, 2, \cdots). \tag{4}$$

For taking $f(x) = x^m$ we see from (3) that $R_n = 0$ when $m < 2n$, while $R_n > 0$

---

* Annales de l'Ecole Normale Supérieure, (3), vol. 1 (1884).

† *Recherches sur les fractions continues*, Annales de la Faculté des Sciences de Toulouse, vol. 7 (1894).

‡ *On the convergence of quadrature formulas between infinite limits*, Bulletin of the Russian Academy of Sciences, 1916 (in Russian).

when $m \geqq 2n$; that is, $L_{m,n}=c_m$ in the first case and $L_{m,n}<c_m$ in the second case.

From now on we shall assume that the function $p(x)$ satisfies the condition

$$\int_\alpha^\beta p(x)dx > 0 \quad \text{if} \quad 0 \leqq \alpha < \beta,$$

and that constants C and R may be found such that

$$\frac{c_m}{(2m+1)!} < CR^{2m} \tag{5}$$

for $m=1, 2, 3, \cdots$.

These conditions being fulfilled, we can establish the convergence of (1) first for a special type of function $f(x)$, namely

$$f(x) = e^{sx^{1/2}} + e^{-sx^{1/2}}$$

denoting by s a complex parameter. For such a choice of $f(x)$ the left member of (1) is

$$\phi_n(s) = \sum_{k=1}^{n} A_k(e^{sx^{1/2}} + e^{-sx^{1/2}}),$$

which, after substituting for exponential functions the corresponding power series, gives

$$\phi_n(s) = 2\sum_{m=0}^{\infty} \frac{L_{m,n}s^{2m}}{2m!}. \tag{6}$$

Now,

$$L_{m,n} \leqq c_m < C(2m+1)!R^{2m},$$

and the series

$$\sum_{m=0}^{\infty}(2m+1)s^{2m}R^{2m}$$

is absolutely and uniformly convergent if

$$|s| \leqq \rho < 1/R. \tag{7}$$

The series (6) is therefore uniformly convergent with respect to n and s provided s satisfies the inequality (7), and since

$$L_{m,n} \to c_m \text{ as } n \to \infty,$$

it follows that

$$\lim \phi_n(s) = 2\sum_{m=0}^{\infty} \frac{c_m s^{2m}}{2m!}$$

uniformly in s. The sum of this series may be represented as a definite integral. This series being convergent for $s=a<1/R$, where a is a positive number, it follows easily that the integral

$$\int_0^\infty p(x)(e^{ax^{1/2}} + e^{-ax^{1/2}})dx$$

exists, and furthermore that

$$2\sum_{m=0}^{\infty} \frac{c_m a^{2m}}{2m!} = \int_0^\infty p(x)(e^{ax^{1/2}} + e^{-ax^{1/2}})dx.$$

This being established, it is not difficult to prove that

$$2\sum_{m=0}^{\infty} \frac{c_m s^{2m}}{2m!} = \int_0^\infty p(x)(e^{sx^{1/2}} + e^{-sx^{1/2}})dx,$$

provided

$$|s| \leqq a < 1/R.$$

That is,

$$\lim \phi_n(s) = \int_0^\infty p(x)(e^{sx^{1/2}} + e^{-sx^{1/2}})dx \tag{8}$$

uniformly in s under the same conditions. The integral in the right member is uniformly convergent if $0 \leqq R(s) \leqq a$, and therefore represents a regular analytic function in the region $0 \leqq R(s) \leqq a$, which we shall denote by $\phi(s)$.

Putting $s=\sigma+i\tau$, it is obvious that

$$|\phi_n(s)| \leqq \phi_n(a),$$

and if $0 \leqq \sigma \leqq a$

$$\phi_n(\sigma) \leqq \phi_n(a) \leqq \sum_{m=0}^{\infty} \frac{c_m a^{2m}}{2m!} = \phi(a).$$

That is,

$$|\phi_n(s)| \leqq \phi(a)$$

whatever n may be if only $0 \leqq R(s) \leqq a$. In the finite region $0 \leqq \sigma \leqq a$, $-T \leqq \tau < T$, the sequence of analytic functions

$$\phi_1(s), \quad \phi_2(s), \quad \phi_3(s), \cdots$$

is therefore *uniformly bounded* and we have established that this sequence converges uniformly to the analytic function $\phi(s)$ in the region $|s| \leqq a$. By a known theorem due to Stieltjes this sequence is *uniformly* convergent to

$\phi(s)$ in the entire region $0 \leqq \sigma \leqq a$, $-T \leqq \tau \leqq T$, that is, (8) holds true in this region.

2. Let a represent again a positive number $<1/R$ and let ρ be any positive number. Starting from the well known formula

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{e^{ws}ds}{s^{1+\rho}} = \frac{w^\rho}{\Gamma(1+\rho)} \text{ if } w > 0,$$

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{e^{ws}ds}{s^{1+\rho}} = 0 \text{ if } w \leqq 0,$$

we find that

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{e^{-st}\phi_n(s)}{s^{1+\rho}}ds = \frac{1}{\Gamma(1+\rho)}\sum_{x_k \geqq t^2} A_k(x_k^{1/2} - t)^\rho \tag{9}$$

for any positive number t, the sum being extended over all x_k satisfying the inequality $x_k \geqq t^2$. We have also

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{e^{-st}\phi(s)}{s^{1+\rho}}ds = \frac{1}{\Gamma(1+\rho)}\int_{t^2}^{\infty} p(x)(x^{1/2} - t)^\rho dx. \tag{10}$$

From the fact that

$$|\phi_n(a + i\tau)| \leqq \phi(a)$$

for every n, and that $\phi_n(a+i\tau)$ converges uniformly to $\phi(a+i\tau)$ in any finite interval $-T \leqq \tau \leqq T$, we can easily infer that

$$\lim_{n=\infty}\int_{a-i\infty}^{a+i\infty}\frac{e^{-st}\phi_n(s)}{s^{1+\rho}}ds = \int_{a-i\infty}^{a+i\infty}\frac{e^{-st}\phi(s)}{s^{1+\rho}}ds,$$

which by (9) and (10) is equivalent to

$$\lim \sum_{x_k \geqq t^2} A_k(x_k^{1/2} - t)^\rho = \int_{t^2}^{\infty} p(x)(x^{1/2} - t)^\rho dx,$$

or, putting $t^2 = \alpha$,

$$\lim \sum_{x_k \geqq \alpha} A_k(x_k^{1/2} - \alpha^{1/2})^\rho = \int_{\alpha}^{\infty} p(x)(x^{1/2} - \alpha^{1/2})^\rho dx. \tag{11}$$

Starting from this important result, we shall prove that

$$\lim \sum_{x_k \geqq \alpha} A_k = \int_{\alpha}^{\infty} p(x)dx. \tag{12}$$

Let ϵ denote any given positive number however small. We choose first a positive number δ so small and another positive number L so large that

$$\int_L^\infty p(x)x^{1/2}dx < \epsilon, \tag{13}$$

$$\frac{1}{1+\delta}\int_0^\alpha p(x)dx - \frac{2\epsilon}{1+\delta} > \int_0^\alpha p(x)dx - 3\epsilon, \tag{14}$$

$$\frac{1}{1-\delta}\int_0^\alpha p(x)dx + \frac{2\epsilon}{1-\delta} < \int_0^\alpha p(x)dx + 3\epsilon. \tag{15}$$

The integrals

$$\int_{\alpha-\eta}^\infty p(x)(x^{1/2} - (\alpha-\eta)^{1/2})^\rho dx \quad \text{and} \quad \int_\alpha^\infty p(x)(x^{1/2} - \alpha^{1/2})^\rho dx$$

being continuous functions of η and ρ in the region $\eta>0$, $\rho>0$, we can take η and σ so small that

$$\int_{\alpha-\eta}^\infty p(x)(x^{1/2} - (\alpha-\eta)^{1/2})^\rho dx < \int_\alpha^\infty p(x)dx + \epsilon, \tag{16}$$

$$\int_\alpha^\infty p(x)(x^{1/2} - \alpha^{1/2})^\rho dx > \int_\alpha^\infty p(x)dx - \epsilon, \tag{17}$$

whenever $0<\rho\leqq\sigma$. Having thus determined L, δ, η, we take ρ so small as to satisfy the inequalities

$$(\alpha^{1/2} - (\alpha-\eta)^{1/2})^\rho > 1-\delta, \quad (4L)^{\rho/2} < 1+\delta, \quad \rho \leqq \sigma, \quad \rho \leqq 1.$$

Finally we can take N so large as to have

$$\sum_{x_k \geqq \alpha-\eta} A_k(x_k^{1/2} - (\alpha-\eta)^{1/2})^\rho < \int_{\alpha-\eta}^\infty p(x)(x^{1/2} - (\alpha-\eta)^{1/2})^\rho dx + \epsilon, \tag{18}$$

$$\sum_{x_k \geqq \alpha} A_k(x_k^{1/2} - \alpha^{1/2})^\rho > \int_\alpha^\infty p(x)(x^{1/2} - \alpha^{1/2})^\rho dx - \epsilon, \tag{19}$$

$$\sum_{x_k \geqq L} A_k(x_k^{1/2} - L^{1/2})^\rho < \int_L^\infty p(x)(x^{1/2} - L^{1/2})^\rho dx + \epsilon, \tag{20}$$

for all $n>N$. Now

$$\sum_{x_k \geqq \alpha-\eta} A_k(x_k^{1/2} - (\alpha-\eta)^{1/2})^\rho \geqq \sum_{x_k \geqq \alpha} A_k(x_k^{1/2} - (\alpha-\eta)^{1/2})^\rho$$

$$\geqq (\alpha^{1/2} - (\alpha-\eta)^{1/2})^\rho \sum_{x_k \geqq \alpha} A_k \geqq (1-\delta)\sum_{x_k \geqq \alpha} A_k,$$

therefore by (18), (16) and (15)

$$\sum_{x_k \geqq \alpha} A_k < \int_\alpha^\infty p(x)dx + 3\epsilon. \tag{21}$$

From (20) and (13) we conclude that

$$\sum_{x_k \geqq 4L} A_k(x_k^{1/2} - L^{1/2})^\rho < 2\epsilon,$$

and for $x_k \geqq 4L$ we have

$$\frac{x_k^{1/2} - \alpha^{1/2}}{x_k^{1/2} - L^{1/2}} < 2\ ;$$

therefore

$$\sum_{x_k \geqq 4L} A_k(x_k^{1/2} - \alpha^{1/2})^\rho < 2^\rho \cdot 2\epsilon < (1 + \delta)2\epsilon.$$

This being established, from (19) and (17) we derive

$$\sum_{\alpha \leqq x_k \leqq 4L} A_k(x_k^{1/2} - \alpha^{1/2})^\rho > \int_\alpha^\infty p(x)dx - 2\epsilon - 2\epsilon(1 + \delta)\ ;$$

but

$$\sum_{\alpha \leqq x_k \leqq 4L} A_k(x_k^{1/2} - \alpha^{1/2})^\rho < (4L)^{\rho/2} \sum_{x_k \geqq \alpha} A_k < (1 + \delta) \sum_{x_k \geqq \alpha} A_k$$

and therefore

$$\sum_{x_k \geqq \alpha} A_k > \int_\alpha^\infty p(x)dx - 5\epsilon. \tag{22}$$

The inequalities (21) and (22) being satisfied, however small ϵ may be, (12) can be considered as established. It follows from (12) that

$$\lim \sum_{\alpha \leqq x_k \leqq \beta} A_k = \int_\alpha^\beta p(x)dx > 0$$

whence we conclude: *for all sufficiently large n there are numbers x_k falling in any given interval (α, β) provided the condition* (5) *is satisfied.*

3. After this important conclusion has been reached, we can apply exactly the same reasoning that Stieltjes uses in an analogous case. Let $f(x)$ denote an integrable function in a finite interval $(0, G)$ and let us suppose $f(x) = 0$ for $x > G$. We shall prove that the formula (1) is convergent for any such function. To this end we introduce a new variable

$$y = \int_0^x p(x)dx$$

increasing with increasing x from 0 to $H=\int_0^G p(x)dx$. The values of this new variable corresponding to $x=x_k$ will be denoted by y_k. We shall also introduce the notation

$$f(x) = \phi(y).$$

The following important inequalities due to Tchebycheff,*

$$A_1 + A_2 + \cdots + A_k > \int_0^{x_k} p(x)dx,$$

$$A_1 + A_2 + \cdots + A_k < \int_0^{x_{k+1}} p(x)dx$$

show that

$$0 \leqq y_1 < A_1 < y_2 < A_1 + A_2 < y_3 < A_1 + A_2 + A_3 < \cdots . \tag{23}$$

The number l being determined by the condition

$$A_1 + A_2 + \cdots + A_{l-1} \leqq H < A_1 + A_2 + \cdots + A_l,$$

we conclude easily from (23) that the sum

$$S = A_1\phi(y_1) + A_2\phi(y_2) + \cdots + A_{l-1}\phi(y_{l-1})$$

converges to the limit

$$\int_0^H \phi(y)dy$$

with indefinitely increasing n provided all A_k approach 0, which is the case, as may be easily seen from the inequality

$$A_k < \int_{x_{k-1}}^{x_{k+1}} p(x)dx$$

combined with the final conclusion in §2. But the sum

$$A_1f(x_1) + A_2f(x_2) + \cdots + A_nf(x_n)$$

can differ from S only by the term $A_l\,\phi(y_l) = A_lf(x_l)$ which converges to 0, that is

$$\lim\,(A_1f(x_1) + A_2f(x_2) + \cdots + A_nf(x_n)) = \int_0^H \phi(y)dy = \int_0^G p(x)f(x)dx$$

as was stated.

---

* Tchebycheff, *Sur les valeurs limites des intégrales*, Journal de Mathématiques, 1874.

4. Now we can prove that the convergence of (1) is assured for any integrable function which for sufficiently large x satisfies the inequality

$$|f(x)| < x^m$$

where m is a certain integer. Let us denote by ϵ a positive number taken arbitrarily and let us determine G by the condition

$$\int_G^\infty p(x)x^m dx < \epsilon.$$

Let $F(x)$ be defined as follows:

$$F(x) = f(x) \quad \text{for} \quad 0 \leqq x \leqq G,$$
$$F(x) = 0 \quad \text{for} \quad x > G.$$

We have, by the results of §3,

$$\lim \sum_{k=1}^{n} A_k F(x_k) = \int_0^\infty p(x)F(x)dx,$$

or

$$\lim \sum_{x_k \leqq G} A_k f(x_k) = \int_0^G p(x)f(x)dx.$$

In particular

$$\lim \sum_{x_k \leqq G} A_k x_k^m = \int_0^G p(x)x^m dx,$$

and as

$$\sum_{k=1}^{n} A_k x^m = \int_0^\infty p(x)x^m dx,$$

we get

$$\lim \sum_{x_k > G} A_k x_k^m = \int_G^\infty p(x)x^m dx < \epsilon,$$

that is

$$\sum_{x_k > G} A_k x_k^m < 2\epsilon$$

for all sufficiently large n, and a fortiori,

$$\left| \sum_{x_k > G} A_k f(x_k) \right| < 2\epsilon.$$

Now, for sufficiently large n,

$$\left| \sum_{x_k \leqq G} A_k f(x_k) - \int_0^\infty p(x)f(x)dx \right| < \epsilon,$$

and therefore

$$\left| \sum_{k=1}^{n} A_k f(x_k) - \int_0^\infty p(x) f(x) dx \right| < 3\epsilon$$

for all sufficiently large n; that is,

$$\lim \sum_{k=1}^{n} A_k f(x_k) = \int_0^\infty p(x) f(x) dx.$$

5. Now we can take the last step and prove the convergence of (1) for for any integrable function satisfying for sufficiently large x the inequality

$$| f(x) | < \frac{e^{rx^{1/2}}}{x^{1+\rho}}$$

where ρ is any positive number <1 and

$$r = 1/R,$$

R being the same number as in (5)

For this purpose we shall first establish a certain inequality. If a_k, $b_k (k=1, 2, 3, \cdots)$ denote positive numbers, we have

$$\sum_1^n a_k b_k \leqq \left(\sum_1^n a_k^\sigma \right)^{1/\rho} \left(\sum_1^n b_k^{\sigma/(\sigma-1)} \right)^{1-1/\sigma},$$

provided $\sigma > 1$. Here we take

$$\sigma = 1/\rho, \; a_k = A_k^\rho x_k^{(m-1)\rho}, \; b_k = A_k^{1-\rho} x_k^{m(1-\rho)},$$

which gives

$$\sum_1^n A_k x_k^{m-\rho} \leqq \left(\sum_1^n A_k x_k^{m-1} \right)^\rho \left(\sum_1^n A_k x_k^m \right)^{1-\rho},$$

whence

$$\sum_1^n A_k x_k^{m-\rho} \leqq c_{m-1} c_m^{1-\rho}. \tag{24}$$

But by (5),

$$c_{m-1} < C(2m-1)! R^{2m-2},$$
$$c_m < C(2m+1)! R^{2m},$$

from which, with (24),

$$\sum_1^n A_k x_k^{m-\rho} < CR^{2m-2\rho}(2m-1) \; [2m(2m+1)]^{1-\rho}, \tag{25}$$

which is the required inequality. Now we consider the function

$$f(x) = \frac{e^{rx^{1/2}} + e^{-rx^{1/2}}}{1 + x^{1+\rho}},$$

and form the sum

$$S_n = \sum_1^n A_k f(x_k) = \sum_1^n A_k \frac{e^{rx_k^{1/2}} + e^{-rx_k^{1/2}}}{1 + x_k^{1+\rho}},$$

which, putting

$$\gamma_m = \sum_1^n \frac{A_k x_k^m}{1 + x_k^{1+\rho}},$$

can be represented by the series

$$S_n = 2 \sum_{m=0}^{\infty} \frac{\gamma_m}{2m!} r^{2m}.$$

As

$$\gamma_m < \sum_1^n A_k x_k^{m-1-\rho},$$

it follows from (25) that

$$\frac{\gamma_m}{2m!} r^{2m} < \frac{C}{R^{2\rho+2}} \frac{1}{2m[(2m-2)(2m-1)]^{\rho}},$$

which shows that the series converges *uniformly in n*.

By what has been established in §4 we have

$$\lim \gamma_m = \int_0^{\infty} \frac{p(x) x^m}{1 + x^{1+\rho}} dx,$$

and therefore

$$\lim S_n = \int_0^{\infty} p(x) \frac{e^{rx^{1/2}} + e^{-rx^{1/2}}}{1 + x^{1+\rho}} dx.$$

We proceed now in the same way as in §4. For any given $\epsilon > 0$, a sufficiently large number G is determined by the condition

$$\int_G^{\infty} p(x) \frac{e^{rx^{1/2}} + e^{-rx^{1/2}}}{1 + x^{1+\rho}} dx < \epsilon.$$

Applying then the result established in §3, we have

$$\lim \sum_{x_k \leqq G} A_k \frac{e^{rx_k^{1/2}} + e^{-rx_k^{1/2}}}{1 + x_k^{1+\rho}} = \int_0^G p(x) \frac{e^{rx^{1/2}} + e^{-rx^{1/2}}}{1 + x^{1+\rho}} dx,$$

and therefore

$$\sum_{x_k>G} A_k \frac{e^{rx_k^{1/2}} + e^{-rx_k^{1/2}}}{1 + x_k^{1+\rho}} < 2\epsilon$$

for sufficiently large n. By hypothesis

$$|f(x)| < \frac{e^{rx^{1/2}}}{x^{1+\rho}} < 2\,\frac{e^{rx^{1/2}} + e^{-rx^{1/2}}}{1 + x^{1+\rho}}$$

as soon as x exceeds a certain finite limit, therefore

$$\left| \sum_{x_k>G} A_k f(x_k) \right| < 4\epsilon$$

for sufficiently large n. On the other hand

$$\left| \sum_{x_k \leqq G} A_k f(x_k) - \int_0^G p(x)f(x)dx \right| < \epsilon$$

for large n, from which, with the preceding, it follows that

$$\left| \sum_{k=1}^{n} A_k f(x_k) - \int_0^\infty p(x)f(x)dx \right| < 6\epsilon$$

for all sufficiently large n. The convergence of (1) for all functions satisfying the condition

$$|f(x)| < \frac{e^{rx^{1/2}}}{x^{1+\rho}}$$

is now completely proved.

6. Suppose we deal with the infinite interval $(a, +\infty)$ where a is any finite number, so that we have

$$\int_a^\infty p(x)f(x)dx = A_1 f(x_1) + A_2 f(x_2) + \cdots + A_n f(x_n)$$

whenever $f(x)$ is a polynomial of degree $\leqq 2n-1$. Putting

$$p_1(t) = p(a + t)\,; \quad t_k = x_k - a$$

we have

$$\int_0^\infty p_1(t)\phi(t)dt = A_1\phi(t_1) + A_2\phi(t_2) + \cdots + A_n\phi(t_n)$$

for any polynomial $\phi(t)$ whose degree does not exceed $2n-1$. According to what has been established, the last formula converges whenever the moments of $p_1(t)$

$$\gamma_m = \int_0^\infty p_1(t)t^m dt$$

satisfy an inequality of the form

$$\gamma_m < D(2m+1)!\, R^{2m} \tag{26}$$

with certain constants D and R, for any function $\phi(t)$ integrable in any finite interval and for large values of t satisfying the condition

$$|\phi(t)| < \frac{e^{r t^{1/2}}}{t^{1+\rho}}$$

where $r = 1/R$ and ρ is any positive number <1. We have

$$\gamma_m = \int_a^\infty p(x)(x-a)^m dx,$$

so that for $a>0$

$$\gamma_m < \int_a^\infty p(x)x^m dx < c_m,$$

while for $a<0$

$$\gamma_m < c_0|a|^m + c_m + \frac{m}{1}c_{m-1}|a| + \frac{m(m-1)}{1\cdot 2}c_{m-2}|a|^2 + \cdots,$$

and, if the condition

$$c_m < C(2m+1)!\, R^{2m}$$

is satisfied,

$$\gamma_m < c_0|a|^m + C(2m+1)!\, R^{2m}\left[1 + \frac{m}{1\cdot 2m(2m+1)}\frac{|a|}{R^2} + \frac{m(m-1)}{1\cdot 2\cdot(2m-2)(2m-1)2m(2m+1)}\left(\frac{|a|}{R^2}\right)^2 + \cdots\right],$$

whence

$$\gamma_m < c_0|a|^m + Ce^{|a|/R^2}(2m+1)!\, R^{2m} < 2Ce^{|a|/R^2}(2m+1)!\, R^{2m}$$

for all sufficiently large m. In every case (26) holds with

$$D = 2Ce^{|a|/R^2}$$

and we can draw the following final conclusion:

The quadrature formula related to the interval (a, ∞) *converges whenever the moments satisfy the condition*

$$c_m < C(2m+1)!R^{2m}$$

with certain constants C *and* R, *for any integrable function satisfying the inequality* (*for sufficiently large* x)

$$|f(x)| < \frac{e^{rx^{1/2}}}{x^{1+\rho}}$$

where $r = 1/R$ *and* ρ *is any positive number* < 1.

Taking

$$f(x) = \frac{1}{z-x}$$

with any complex z not belonging to the interval $(a, +\infty)$, we see that the continued fraction

$$\int_a^\infty \frac{p(x)dx}{z-x} = \frac{\lambda_0}{z-\alpha_0} - \frac{\lambda_1}{z-\alpha_1} - \cdots$$

converges and represents the integral if only the moments c_m satisfy the above mentioned condition.

To show how far-reaching this criterion is, let us take $a=0$ and $p(x)=e^{-x^{1/2}}$.

In this case $c_m=(2m+1)!$, so that $R=1$, $C=1$. The quadrature formula converges for any function satisfying the inequality

$$|f(x)| < \frac{e^{x^{1/2}}}{x^{1+\rho}}$$

for large x and, in particular, the continued fraction converges and represents the integral. We obtain the same result for

$$p(x) = e^{-x^\lambda}$$

with $\lambda > \frac{1}{2}$ because the c_m corresponding to this $p(x)$ are obviously less than in the preceding case. But the convergence ceases with $\lambda = \frac{1}{2}$, for it is known that for $\lambda < \frac{1}{2}$ the problem of moments becomes indeterminate.

7. It remains to consider the case of an infinite interval $(-\infty, \infty)$. The analysis of this case is but slightly different, and we can confine ourselves to a very brief outline. We retain the fundamental assumption that

$$\int_\alpha^\beta p(x)dx > 0$$

for any two numbers α, β satisfying the condition $\alpha<\beta$. As to the moments, we consider only those of even order and assume the existence of two constants C and R such that

$$c_{2m} < C(2m)!\,R^{2m}.$$

For moments of odd order we have

$$|c_{2m+1}| \leqq (c_{2m}c_{2m+2})^{1/2} < C(2m+1)!\,R^{2m+1}\left(\frac{2m+2}{2m+1}\right)^{1/2} < 2C(2m+1)!\,R^{2m+1}.$$

By means of these inequalities it can be shown in exactly the same way as in §1 that the analytic function

$$\phi_n(s) = \sum_{k=1}^{n} A_k e^{-sx_k}$$

converges uniformly in s to

$$\phi(s) = \int_{-\infty}^{\infty} p(x)e^{-sx}dx$$

provided

$$|s| \leqq \rho < 1/R.$$

The function $\phi(s)$ is a regular analytic function in the region $-a \leqq R(s) \leqq a$ where $0<a<1/R$.

In the same region the sequence

$$\phi_1(s), \quad \phi_2(s), \quad \phi_3 s), \cdots$$

is uniformly bounded and by Stieltjes' theorem $\phi_n(s)$ converges to $\phi(s)$ *uniformly* in a finite region

$$-a \leqq \sigma \leqq a, \quad -T \leqq \tau \leqq T.$$

This being established, we reach in the same way as in §2 the important result

$$\lim \sum_{x_k \leqq t} A_k(t-x_k)^\rho = \int_{-\infty}^{t} p(x)(t-x)^\rho dx$$

for every $\rho>0$, whence we can conclude that

$$\lim \sum_{x_k \leqq a} A_k = \int_{-\infty}^{a} p(x)dx,$$

and

$$\lim \sum_{a<x_k \leqq \beta} A_k = \int_{a}^{\beta} p(x)dx > 0.$$

This shows that for sufficiently large n there always numbers x_k belonging to any given interval (α, β).

Again repeating the same reasoning as in §3 we reach the conclusion that formula (1) converges for any integrable function which is zero outside a given finite interval. Then, as in §4, this conclusion is extended to any function integrable in any finite interval and for sufficiently large x satisfying the inequality

$$|f(x)| < x^m.$$

Finally, the same conclusion may be extended to functions which for large x satisfy the inequality

$$|f(x)| < \frac{e^{r|x|}}{|x|^{1+\rho}}$$

where $r = 1/R$ and ρ is a positive number <1.

To prove this statement we observe that, the convergence being once established for a *positive* function $F(x)$, it is secured for any function satisfying the condition

$$|f(x)| < F(x)$$

at least for sufficiently large x. This may be easily proved by the same reasoning as developed in §4. We take

$$F(x) = \frac{e^{rx} + e^{-rx}}{1 + |x|^{1+\rho}},$$

where $0 < \rho < 1$, and obtain

$$S = \sum_1^n A_k F(x_k) = 2 \sum_{m=0}^{\infty} \frac{\gamma_m}{2m!} r^{2m},$$

γ_m being defined by

$$\gamma_m = \sum_1^n A_k \frac{{x_k}^{2m}}{1 + |x_k|^{1+\rho}}.$$

We have obviously

$$\gamma_m < \sum_1^n A_k |x_k|^{2m-1-\rho},$$

but

$$\sum_1^n A_k |x_k|^{2m-1-\rho} \leqq \left(\sum_1^n A_k x_k^{2m-2} \right)^{\rho} \left(\sum_1^n A_k |x_k|^{2m-1} \right)^{1-\rho},$$

or

$$\sum_1^n A_k |x_k|^{2m-1-\rho} \leqq 2CR^{2m-1-\rho} \frac{(2m-1)!}{(2m-1)^\rho}.$$

The series

$$\sum_{m=0}^{\infty} \frac{\gamma_m}{2m!} r^{2m}$$

is therefore uniformly convergent in n, and as

$$\lim \gamma_m = \int_{-\infty}^{\infty} p(x) \frac{x^{2m}}{1+|x|^{1+\rho}} dx,$$

we get

$$\lim S = \int_{-\infty}^{\infty} p(x) \frac{e^{rx}+e^{-rx}}{1+|x|^{1+\rho}} dx,$$

so that the convergence is established for $F(x)$.

Now if

$$|f(x)| < \frac{e^{r|x|}}{|x|^{1+\rho}}$$

for sufficiently large x, it is obvious that under the same condition

$$|f(x)| < 2F(x),$$

which proves our statement.

8. In cases in which the rapidity of increase of the moments is less than was assumed, the class of functions for which the convergence is secured may be extended. Let us take for example $p(x)=e^{-x^2}$ and the interval $(-\infty, \infty)$. In this case

$$c_{2m} = \pi^{1/2} \frac{2m!}{2^{2m}m!}.$$

Taking

$$F(x) = \frac{e^{x^\rho}}{1+|x|^{1+\rho}}, \quad 1 > \rho > 0,$$

we have

$$S = \sum_1^n A_k F(x_k) = \sum_{m=0}^{\infty} \frac{\gamma_m}{m!},$$

where

$$\gamma_m = \sum_1^n A_k \frac{x_k^{2m}}{1+|x_k|^{1+\rho}}.$$

As

$$\gamma_m < \sum_1^n A_k \mid x \mid^{2m-1-\rho}$$

and

$$\sum_1^n A_k \mid x_k \mid^{2m-1-\rho} \leqq \left(\sum_1^n A_k x_k^{2m-2} \right)^{\rho} \left(\sum_1^n A_k \mid x_k \mid^{2m-1} \right)^{1-\rho}$$

$$< \frac{\pi^{1/2} \cdot 2^{(1+\rho)/2}}{2^{2m-1}} \frac{(2m-2)!}{(m-1)!} (2m-1)^{(1-\rho)/2},$$

it is easy to see that the series representing S converges uniformly in n and therefore

$$\lim S = \int_{-\infty}^{\infty} \frac{e^{x^2}}{1 + \mid x \mid^{1+\rho}} dx.$$

That is, the quadrature formula corresponding to $p(x) = e^{-x^2}$ converges for $F(x)$ and a fortiori it will be convergent for any function satisfying the condition

$$\mid f(x) \mid < \frac{e^{x^2}}{\mid x \mid^{1+\rho}}, \quad \rho > 0,$$

for sufficiently large x

Another example of the same kind is given by the formula corresponding to $p(x) = e^{-x} x^{\alpha}$ and the interval $(0, \infty)$. In this case the convergence is secured for any integrable function satisfying the condition

$$\mid f(x) \mid < \frac{e^{x}}{x^{\alpha+1+\rho}}, \quad \rho > 0,$$

for sufficiently large x.

CARLETON COLLEGE,
NORTHFIELD, MINN.

SECOND-ORDER LINEAR SYSTEMS WITH SUMMABLE COEFFICIENTS*

BY

J. H. STURDIVANT

Sturm and his successors† have examined in detail properties of the zeros of solutions of linear differential systems similar to (1), (3) below, and of linear combinations similar to (2). They have commonly employed the continuity and existence of derivatives of the coefficients. Under more lenient assumptions, the writer will obtain sufficient conditions that the zeros of the linear combination (2) shall be monotonic functions of μ, the parameter. From this fundamental theorem, one may develop, with the less restrictive hypotheses, oscillation and existence theorems which extend many results obtained in the past for classical systems.‡

We shall consider the linear differential system

$$(1) \qquad \begin{aligned} y_x'(x,\mu) &= K(x,\mu)z(x,\mu), \\ z_x'(x,\mu) &= G(x,\mu)y(x,\mu), \end{aligned}$$

and the linear combination

$$(2) \qquad \phi(x,\mu) = \alpha(x,\mu)z(x,\mu) - \beta(x,\mu)y(x,\mu).$$

The following conditions shall be imposed:

Conditions S:

(a) K and G are summable§ in x on X: $a \leqq x \leqq b$, for every μ on M: $M_1 < \mu < M_2$. $|K(x,\mu)|$, $|G(x,\mu)| \leqq U(x)$.‖

(b) α, β are absolutely continuous in x on X for any fixed μ on M and are continuous in μ for fixed x on XM, $|\alpha_x'|$, $|\beta_x'| \leqq U(x)$ on XM.

(c) K, G, α_x', and β_x' are continuous in μ for every x, on XM.

(d) $\alpha \neq 0$ on¶ XM.

---

* Presented to the Society, September 9, 1926; received by the editors in May, 1927.

† For references to the literature, see Ettlinger, these Transactions, vol. 19, p. 79.

‡ The writer gratefully acknowledges Professor Ettlinger's helpful encouragement during the preparation of this paper.

§ Throughout this paper, *summable* is used in the sense of Lebesgue.

‖ The symbol $U(x)$ signifies *some positive, summable function of* x.

¶ We may actually admit, also, $\alpha \equiv 0$ on XM, for in this case condition (g) of S requires $\beta \neq 0$, and the zeros of y and ϕ are coincident. We may then utilize a proof of Theorems B and II, stated in terms of y instead of ϕ.

(e) $K(x, \mu) > 0$ on* X_0M.

(f) $|y| + |z| \neq 0$ on XM.

(g) $\{\bar{\beta}\} \equiv (1/\alpha^2)\{\alpha\beta\} \equiv (1/\alpha^2)[\alpha\beta_x' - \alpha_x'\beta + \beta^2 K - \alpha^2 G] > c > 0$ on X_0M (or $< c < 0$ on X_0M).

Concerning the system composed of (1) with the initial conditions

$$y(a, \mu) = A(\mu), \quad z(a, \mu) = B(\mu), \tag{3}$$

where A and B are continuous in μ on M, we have the following

THEOREM A. *There exists a unique pair of functions* (y, z), *continuous in* (x, μ) *and absolutely continuous in* x *on* X *for* μ *fixed on* M, *which satisfy* (1) *on* X_0M *and* (3) *on* M.†

Such a pair of functions, (y, z), constitutes a *solution* of the system (1), (3).

An implicit function theorem due to Ettlinger‡ (and hereafter referred to as Theorem B′) will be employed in a modified form, as follows:

THEOREM B. *Hypotheses*:

(1) $f(\bar{x}, \bar{\mu}) = 0$, *where* $\bar{x}$ *represents a set of values,* $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$, *on* X: $a \leqq x \leqq b$.

(2) $f(x, \mu)$ *is absolutely continuous in* x *on* X *for any* μ *on* $\overline{M}$: $\overline{M}_1 \leqq \mu \leqq \overline{M}_2$. $f(x, \mu)$ *is continuous in* μ *for fixed* x *on* XM.

(3) $f_x'(x, \mu)$ *is continuous in* μ *on* $\overline{M}$ *for any fixed* x *on* X_0. $|f_x'| \leqq U(x)$ *on* $X\overline{M}$.

(4) $f_x'(x, \mu) > 0$ *or* $f_x'(x, \mu) < 0$ *for almost every* x *on an interval* $|x - p| \leqq k$, $k > 0$, *when* μ *is fixed on* $|\mu - m| \leqq h$, $h > 0$, *and whenever* (p, m) *is a zero of* f *on* $X\overline{M}$. *The numbers* k *and* h *may depend upon the particular* m.

Conclusion: *The function* $x = x(\mu)$, *such that* $f[x(\mu), \mu] \equiv 0$, *consists of a finite number of continuous and single-valued branches on* $X\overline{M}$.

Consider first the case $n = 1$. By Theorem B′, we have a continuous and single-valued function, $x = x(\mu)$, in and on the boundary of $D'F'$, which last is a neighborhood of the point $(\bar{x}, \bar{\mu})$ and is assumed wholly within $X\overline{M}$.

---

* X_0M is to be read "at any fixed μ on M, for every x on X except possibly a set of measure zero."

† Tamarkin, *Concerning some General Problems in the Theory of Ordinary Linear Differential Equations*, etc., Petrograd, 1917, p. 29, ff., and Carathéodory, *Vorlesungen über reelle Funktionen*, Leipzig, 1918, p. 639, and p. 678. See also Ettlinger, Bulletin of the American Mathematical Society, vol. 33, pp. 37–38

‡ *Existence theorems for implicit functions*, Bulletin of the American Mathematical Society, vol. 34, p. 316.

About the point $[x(\bar{\mu}+h), \bar{\mu}+h]$, on the boundary of $D'F'$, we may find a neighborhood into which and on the boundary of which $x(\mu)$ is continued, since all necessary hypotheses hold everywhere within $X\overline{M}$. The function is therefore continuous and single-valued to some μ outside of $D'F'$. We may continue to extend the definition of the function from $\bar{\mu}$ toward the right up to any other μ on $\overline{M}$, obtaining an arc A. Due to a diminishing in the dimensions of the successively chosen neighborhoods, we may not be able to attain every μ on $\overline{M}$. Let $\mu=\mu^*$ be the lower bound of all values we cannot reach. We can approach arbitrarily close to μ^*, and there is at least one point, $P:(x^*, \mu^*)$, which is a limit point of the set of zeros A. It follows, now, from a theorem which Ettlinger† has isolated from Carathéodory, that f is a continuous function of (x, μ) when x and μ vary independently on $X\overline{M}$; the zeros of f, accordingly, form a closed set in the same region. Then $f(x^*, \mu^*)=0$, and we may now obtain a neighborhood of P in which $x=x(\mu)$ is single-valued and continuous. Since the set $A+P$ of zeros of f is closed, the Heine-Borel‡ Theorem allows us to select a finite number of the covering neighborhoods to cover the set, and thence to obtain a finite number of overlapping elements of the continuous, single-valued curve, $x=x(\mu)$. Then this function is continuous and single-valued from $(\bar{x}, \bar{\mu})$ to (x^*, μ^*), where μ^* is any μ on $\overline{M}$.

Permit n to be any integer, as assumed in Hypothesis (1). That this n, the number of zeros for μ fixed at some $\bar{\mu}$, always remains finite is an immediate consequence of the fourth hypothesis. For, if we assume an infinite set of zeros for any fixed $\bar{\mu}$, this set has a limit point c, and $f(c, \bar{\mu})=0$. But almost everywhere on some interval containing c, $f_x'(x, \bar{\mu})>0$ or <0; therefore $(c, \bar{\mu})$ must be an isolated zero, and there is a contradiction. About each point $(\bar{x}_i, \bar{\mu})$ we may draw a rectangular neighborhood R_i, and may trim each neighborhood in the x direction, so that there is no overlapping in that direction. We may now construct for each of the finite number of zeros, just as was done for the single zero above, a function $x_i(\mu)$, continuous and single-valued on $\overline{M}$. None of the branches thus formed will intersect another on $\overline{M}$, for by Theorem B$_1'$ the distance between any two branches has a positive lower bound. In this case, therefore, the function $x(\mu)$ consists of a finite number of single-valued, continuous functions on $\overline{M}$.

With regard to a solution y of the system (1), (3), S, we note immediately that the first three hypotheses of Theorem B are satisfied. Hypothesis (4) follows from conditions (e) and (f) of S, the continuity of z, and the first

† See footnote to Theorem A.

‡ Cf. Hahn, *Theorie der reellen Funktionen*, Berlin, 1921, vol. 1, p. 89.

equation of (1). Therefore, the function $x=x(\mu)$, representing the zeros of y, consists of a finite number of continuous, single-valued functions on M.

Consider $\psi=\phi/\alpha$, where ϕ is defined by (1), (2), and S. In the case of ψ, Hypotheses (2) and (3) of Theorem B are obviously valid for any x on X, if μ is on any closed subinterval of M. We obtain the properties required in Hypothesis (4) through a

LEMMA. *If K is bounded on X for any fixed μ on M, and $\psi(\bar{x}, \bar{\mu})=0$, then $\psi_x'(x, \bar{\mu})>g>0$ (or $<g<0$) almost everywhere on $|x-\bar{x}|\leqq k$, $k>0$, where g is some constant.*

Eliminating z between

$$\psi = z - \bar{\beta}y, \qquad \bar{\beta} = \beta/\alpha,$$

and

$$\psi_x' = -\bar{\beta}Kz + y(G - \bar{\beta}_x')$$

(which latter exists on X_0M), we obtain

$$-\{\bar{\beta}\}y = K\bar{\beta}\psi - \psi_x' \qquad \text{on } X_0M,$$

where

$$\{\bar{\beta}\} = \bar{\beta}_x' + K\bar{\beta}^2 - G.$$

Since $|y|+|z|\neq 0$ on XM, and since y is continuous, we have either $y(x, \bar{\mu})<0$ or $y(x, \bar{\mu})>0$ on an interval F: $|x-\bar{x}|\leqq k^*$. For convenience, let it be negative on F. Actually, $y(x, \bar{\mu})$ has an extremum on F, and we may write $y(x, \bar{\mu})<-h<0$, h being some constant. Given any positive η, there is, moreover, a sub-neighborhood, F', of F such that $|\psi|\leqq\eta$ on F'. Therefore, on setting $|K(x, \bar{\mu})|<K_0$ and $|\bar{\beta}(x, \bar{\mu})|<\bar{\beta}_0$ on F',

$$\psi_x'(x, \bar{\mu}) > \{\bar{\beta}\}h - K_0\eta\bar{\beta}_0,$$

and hence, from condition (g) of S,

$$\psi_x'(x, \bar{\mu}) > ch - K_0\eta\bar{\beta}_0 = g > 0 \text{ on } F_0',$$

if η is sufficiently small.

The cases where $y>0$ or $\{\bar{\beta}\}<c<0$ are treated in the same fashion.

But the zeros of ϕ and of ψ are identical, and we may consequently assert that $x=x(\mu)$, the function such that $\phi[x(\mu), \mu]\equiv 0$, consists of a finite number of continuous, single-valued functions on any closed subinterval of M.

Comparatively simple conditions suffice to render the zeros of ϕ monotonic decreasing functions of μ. They are embodied in Theorem II below, in preparation for which we shall prove Theorem I.†

† Notation: All increments written δf are changes in $f(x, \mu)$ due to change of μ alone. For brevity, only positive $\delta\mu$'s are employed.

THEOREM I. *Hypotheses*:

(1) (y, z) *is a solution of the system* (1), (3), S.

(2) *Either* $A \equiv 0$ *or* $A \neq 0$ *on* M.

(3) *For* $\delta\mu > 0$, $-\delta K, \delta G \leqq 0$, *and if* $A \neq 0$, $\delta(B/A) \leqq 0$. *The three equalities do not hold simultaneously anywhere on* XM. *If* $A \equiv 0$, $|\delta K| + |\delta G| \neq 0$ *at any point of* XM.

(4) $V(y, z) = y\delta z - z\delta y$.

Conclusion: $V(y, z) < 0$ *on* XM.

The derivative of V is

$$V_x' = y_x'\delta z + y\delta z_x' - z_x'\delta y - z\delta y_x'.$$

By means of substitutions from (1), it may be written in a form given essentially by Sturm:*

$$V_x' = \delta G(y^2 + y\delta y) - \delta K \frac{y_x'^2 + y_x'\delta y_x'}{K(K + \delta K)}.$$

Since y and y_x' are continuous in μ, given any $\eta > 0$, there exists an $\epsilon_\eta > 0$ such that whenever $\delta\mu \leqq \epsilon_\eta$, we have $|\delta y|$, $|\delta y_x'| \leqq \eta$. Since, also, $|y| + |y_x'| > 0$ almost everywhere, we may take $\delta\mu$ small enough that

$$\operatorname{sgn} V_x' = \operatorname{sgn}\left[y^2\delta G - y_x'^2 \frac{\delta K}{K(K + \delta K)}\right] = -1 \text{ on } XM,$$

save where $\delta G = \delta K = 0$. We have $V(a, \mu) = A\delta B - B\delta A$, and

$$\delta(B/A) = \frac{(A\delta B - B\delta A)}{A^2}\left(1 - \frac{\delta A}{A + \delta A}\right).$$

For $\delta\mu$ small, again, and $A \neq 0$,

$$\operatorname{sgn} \delta(B/A) = \operatorname{sgn} V(a),$$

and $V(a, \mu) \leqq 0$ on M. If $A \equiv 0$, $V(a) \equiv 0$, and need not be considered. The conclusion of the theorem follows at once from the relation

$$V(x, \mu) = \int_a^x V_t'(t, \mu)dt + V(a, \mu).$$

THEOREM II. *Hypotheses*:

(1) $\phi(y, z) = \alpha(x, \mu)\, z(x, \mu) - \beta(x, \mu)\, y(x, \mu)$.

---

* Bôcher, *Leçons sur les Méthodes de Sturm*, Paris, 1917, p. 54.

(2) *y and z satisfy the system* (1), (3), S *on* X_0M, *with* $\{\bar\beta\}>c>0$, *and* ϕ *satisfies the conditions of* S. *K is bounded on X for any fixed μ on M.*

(3) $\delta(B/A)\leqq 0$ *on M if* $A\neq 0$; $-\delta K$, δG, *and* $-\delta(\beta/\alpha)\leqq 0$ *for any x on X, when μ increases on M. Not all of the equalities hold simultaneously.*

Conclusion: The zeros of ϕ move to the left with increase of μ; i.e., if $\phi[x_i(\mu), \mu]\equiv 0$, $\delta x_i<0$ *on M.*

Let $\psi=\phi/\alpha=z-\bar\beta y$, $\bar\beta=\beta/\alpha$. Consider the function

$$V(\psi,\psi_x') = \psi\delta\psi_x' - \psi_x'\delta\psi.$$

Simplification by means of (1) gives us

$$(4)\quad V(\psi,\psi_x') = V(y,z)\{\bar\beta\} + y^2(G\delta\bar\beta - \bar\beta\delta G + \bar\beta\delta\bar\beta_x' - \bar\beta_x'\delta\bar\beta) + z^2(-\bar\beta\delta K - K\delta\bar\beta) + yz(\delta G - \delta\bar\beta_x' + \bar\beta^2\delta K) + O(\delta\mu^2).$$

Here $O(\delta\mu^2)$ denotes additional terms of the second order in $\delta\mu$, such as $z\delta\bar\beta\delta y$, which, because of continuity, we can render as small as desired relative to first-order terms in $\delta\mu$ by choosing $|\delta\mu|$ sufficiently small.

Let y and z be solved for in terms of ψ and ψ_x'; the result is

$$y = \frac{K\bar\beta\psi + \psi_x'}{-\{\bar\beta\}}, \quad z = \frac{\psi(G-\bar\beta_x') + \bar\beta\psi_x'}{-\{\bar\beta\}}.$$

Since ψ is continuous in μ, if we let $\bar x=x_i(\bar\mu)$ represent a zero of ψ, then given $\epsilon>0$, there exists $\eta_\epsilon>0$ such that whenever $|\mu-\bar\mu|<\eta_\epsilon$, $|E_1|$, $|E_2|<\epsilon$, where we write

$$y(\bar x,\mu) = \frac{\psi_x'(\bar x,\mu)}{-\{\bar\beta\}} + E_1, \quad z(\bar x,\mu) = \frac{\bar\beta\psi_x'(\bar x,\mu)}{-\{\bar\beta\}} + E_2.$$

Let these values be substituted into (4). On the same neighborhood, we find

$$V[\psi(\bar x,\mu),\psi_x'(\bar x,\mu)] = [V(y,z)\{\bar\beta\} - (1/\{\bar\beta\})\psi_x'^2(x,\mu)\delta\bar\beta + O(\delta\mu^2)]_{x=\bar x}.$$

But $\{\bar\beta\}>0$. Also, $-\delta\bar\beta=-\delta(\beta/\alpha)\leqq 0$, and $V(y,z)<0$ on XM. Therefore, on an interval D: $|\mu-\bar\mu|<\eta$, where $\eta\leqq\eta_\epsilon$,

$$V[\psi(\bar x,\mu),\psi_x'(\bar x,\mu)] < 0.$$

If $x_i=x_i(\mu)$ represents the ith zero of ψ on X for a certain μ, and if $x_1=x_i(\mu_1)$, then the preceding inequality shows that in a neighborhood D_i: $|\mu-\mu_1|<\eta$,

$$(5)\qquad -\psi_x'(x_1,\mu)\delta\psi(x_1,\mu) < 0.$$

We set $x_2=x_i(\mu_2)=x_1+\delta x_i$, where $\delta\mu=\mu_2-\mu_1<\eta$, and may write

$$\int_{x_1}^{x_1+\delta x_i} \psi_i'(t,\mu_1)dt = \psi(x_2,\mu_1). \tag{6}$$

Since $x_i(\mu)$ is continuous, by taking $\delta\mu$ small enough δx_i may be made smaller than the distance between two consecutive zeros of $\psi(x, \mu_1)$ or $\psi(x, \mu_2)$. Then

$$\begin{aligned} \operatorname{sgn}\psi(x_1,\mu_2) &= -\operatorname{sgn}\psi(x_2,\mu_1) \\ &= \operatorname{sgn}\delta\psi(x_1,\mu). \end{aligned} \tag{7}$$

The possibility that $\delta x_i = 0$ is eliminated as an immediate consequence of a relation following from (5), namely,

$$\delta\psi(x_1,\mu) = \psi(x_1,\mu_2) \neq 0.$$

From the Lemma, and from (5) and (7), it follows that in (6),

$$\delta x_i < 0.$$

The zeros of ϕ are identical with those of ψ, and hence they also move toward the left on X with increase of μ on M.

By an argument very similar to the preceding one, employing $V(y, z)$ instead of $V(\psi, \psi_x')$, we may show independently of Theorem II that the zeros of y (or of z) move to the left on X as μ increases on M. When $\alpha \equiv 0$ on XM, the zeros of ϕ and of y are identical; hence the conclusion of Theorem II is valid for $\alpha \equiv 0$ as well as for $\alpha \neq 0$ on XM.

The writer has employed Theorem II to obtain existence and oscillation theorems for the Sturmian and for the general self-adjoint systems under the hypotheses of S concerning continuity and existence of derivatives. The methods follow, in general, those of Bôcher* and of Ettlinger,† and the results of Ettlinger concerning the distribution of characteristic numbers, etc., may be duplicated.

---

* Bôcher, loc. cit.

† Ettlinger, these Transactions, vol. 19, pp. 79–96, and vol. 22, pp. 136–143.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

CONCERNING THE ARC-CURVES AND BASIC SETS OF A CONTINUOUS CURVE*

BY

W. L. AYRES

In this paper we define and discuss some of the properties of certain subsets of a continuous curve,† which we call the arc-curves. Probably the most interesting property of these subsets is that whenever the arc-curve is closed, it is itself a continuous curve. The arc-curves are intimately connected with the theory of separation of points in continuous curves, since if a closed set separates a pair of points in the arc-curve of the pair of points with respect to a given continuous curve, the closed set separates the points in the given continuous curve. We define a basic set of a continuous curve as any set whose arc-curve is the continuous curve itself. Necessary and sufficient conditions are given in order that a subset of a continuous curve be a basic set and in order that a basic set be irreducible.

It is assumed that all point sets of the paper lie in a two-dimensional euclidean space.‡

Notation. Wherever a symbol H is used to denote a point set, the symbol $\overline{H}$ denotes the point set consisting of the points of H plus all limit points of H. If u and v are points of a simple continuous arc α, the symbol uv denotes the subarc of α with u and v as end points if $u \neq v$, and if $u = v$, the symbol uv denotes the single point $u(=v)$. Hereafter we shall use the term "arc" to mean "simple continuous arc." If uv is an arc with end points u and v, the symbols $<uv, uv>$, $<uv>$ denote $uv-u$, $uv-v$, $uv-u-v$ respectively. If A and B are sets of points, the symbol $A \subset B$ means A is a subset of B, but it is *not* implied that A is a *proper* subset of B.

I

LEMMA. *If P is a point of a set K such that $K-P$ is the sum of two mutually separated point sets K_1 and K_2 and A and B are distinct points of K_1+P, then K_2 contains no point of any arc from A to B which belongs entirely to K.*

* Presented to the Society, May 7, 1927; received by the editors in July, 1927.

† A *continuous curve* is a point set which is connected and connected im kleinen. This is sometimes called a generalized continuous curve. See S. Mazurkiewicz, *Sur les lignes de Jordan*, Fundamenta Mathematicae, vol. 1 (1920), p. 193.

‡ (Added in proof.) In a paper *Concerning continuous curves in space of n dimensions*, which is not yet published, I have recently shown that all the results of this paper, except part of Theorem 10, hold in euclidean space of n dimensions.

Definition. Let K be a subset of a continuous curve M. The set of all points $[P]$, such that P lies on some arc of M whose end points are points of K, will be called the *arc-curve of K with respect to M* and will be denoted by $M(K)$. In case K consists of a single point, we define the arc-curve of K with respect to M to be this single point.

It is evident that the set $M(K)$ is connected and contains the set K. We shall proceed to determine further properties of the point set $M(K)$.

THEOREM 1. *If H and K are subsets of the continuous curve M and H is a subset of K, then $M(H)$ is a subset of $M(K)$.*

The truth of this theorem follows easily from the definition.

THEOREM 2. *If K is any subset of the continuous curve M, the point set $\overline{M(K)}-M(K)$ is a subset of the point set $\overline{K}-K$.*

Suppose there exists a point x which belongs to $\overline{M(K)}-M(K)$ but not to $\overline{K}-K$. Then x belongs to $M-\overline{K}$. Let y be a point of K, and let xy denote an arc of M whose end points are x and y. If d is any component* of $M-xy$, let P_{xd} and P_{yd} denote the first and last limit points of d on xy in the order from x to y. The points P_{xd} and P_{yd} exist (but are not necessarily distinct) for each component d since the set of limit points of d on the arc xy is closed. Let z_1 denote the first point of $\overline{K}+y$ on the arc xy in the order x to y. Evidently $x \neq z_1$. If K consists of a single point or if $M-xy$ is vacuous, we may arrive at a contradiction easily by using the Lemma. If neither of these two hold, let S be the set of all points $[P_{xd}]$ such that d contains a point of K. The arc xy contains a point z_2 such that either (a) z_2 belongs to S and every point of S lies on the subarc z_2y of xy, or (b) z_2 belongs to $xy-S$, every point of S lies on the subarc z_2y of xy and z_2 is a limit point of S, or (c) S is vacuous and $z_2=z_1$. Since x does not belong to $M(K)$, x is not a point of S. If x is a limit point of S, it is also a limit point of K† which is contrary to the hypothesis. Then $z_2 \neq x$. Let z_3 be the first point of the set z_1+z_2 on the arc xy in the order x to y. If $z_3=y$, we may obtain a contradiction by using the Lemma. Then $x \neq z_3 \neq y$.

Suppose $<xz_3$ contains a point p such that $M-xy$ contains no component d such that p lies between P_{xd} and P_{yd}. If x and y lie in the same

---

* A set K is said to be a *component* of a set H, if K is a connected subset of H and there is no connected subset of H which contains K as a proper subset.

† Since only a finite number of the components of $M-xy$ are of diameter greater than any given positive number. Cf. my paper, *Note on a theorem concerning continuous curves*, Annals of Mathematics, (2), vol. 28 (1927), pp. 501–2.

component of $M-p$, there exists an arc α with end points x and y and lying entirely in the set $M-p$.* In the order from x to y the arc α has a last point t_1 in common with the subarc xp of the arc xy; and on the subarc t_1y of α there is a first point t_2 in common with the subarc py of xy. If d_α denotes the component of $M-xy$ containing the subarc t_1t_2 of α, the point p lies between the points P_{zd_α} and P_{yd_α} contrary to our supposition that no such component exists. Hence

$$M - p = M_x + M_y,$$

where M_x and M_y are non-vacuous mutually separated sets containing x and y respectively, and M_x is connected. Then $M_y+p \supset K$, and by the Lemma, x is not a limit point of $M(K)$, which contradicts the hypothesis of the theorem. Thus we have shown that if p is any point of $<xz_3$ there is a set d_p such that p lies between P_{zd_p} and P_{yd_p}.

Since only a finite number of the components of $M-xy$ are of diameter greater than a given positive number, there is a component d_1 such that (a) z_3 lies between P_{zd_1} and P_{yd_1} on the arc xy, (b) if d is any component of $M-xy$ such that z_3 lies between P_{zd} and P_{yd} on the arc xy, then P_{zd} lies on the subarc yP_{zd_1}. If $P_{zd_1} \neq x$, there is a component d_2 of $M-xy$ such that (a) P_{zd_1} lies between P_{zd_2} and P_{yd_2}, (b) if d is any component of $M-xy$ such that P_{zd_1} lies between P_{zd} and P_{yd}, then P_{zd} lies on the subarc yP_{zd_2} of the arc xy. Continue this process. In general, if $P_{zd_{i-1}} \neq x$, there is a component $d_i (i>1)$ of $M-xy$ such that (a) $P_{zd_{i-1}}$ lies between P_{zd_i} and P_{yd_i} on the arc xy, (b) if d is any component of $M-xy$ such that $P_{zd_{i-1}}$ lies between P_{zd} and P_{yd}, then P_{zd} lies on the subarc yP_{zd_i} of the arc xy. If for any value of i, $P_{zd_i}=x$, the process terminates at this point; if not, we continue the process indefinitely.

In the set $z_3P_{yd_1}>$ there is a point z_0 such that either (1) there is a component d_0 of $M-xy$ containing a point z of K such that $P_{zd_0}=z_0$, or (2) z_0 is a point of K. In case (1), by a theorem due to R. L. Wilder,† there exists an arc A_0 whose end points are z and z_0 and such that A_0 belongs to d_0 except for the point z_0. In case (2), let A_0 and z each denote the single point z_0. For every $i>0$ such that d_i is defined, there exists an arc A_i with end points P_{zd_i} and P_{yd_i} and such that d_i contains $<A_i>$.

If there exists an integer n such that $P_{zd_n}=x$ and n is odd, let

---

* Cf. R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), p. 255, Theorem 1.

† *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), Theorem 1, p. 342.

$$\alpha_1 = A_0 + z_0P_{yd_2} + A_2 + P_{xd_2}P_{yd_4} + A_4 + \cdots + P_{xd_{2i-2}}P_{yd_{2i}} + A_{2i} + \cdots + A_{n-1} + P_{xd_{n+1}}x,$$

$$\alpha_2 = yP_{yd_1} + A_1 + P_{xd_1}P_{yd_3} + A_3 + \cdots + P_{xd_{2i-1}}P_{yd_{2i+1}} + A_{2i+1} + \cdots + P_{xd_{n-2}}P_{yd_n} + A_n.$$

It is evident that α_1 and α_2 are arcs with end points z and x and y and x respectively, and that they have only the point x in common. Then $\alpha_1+\alpha_2$ is an arc of M from y to z containing the point x. If there exists an integer n such that $P_{xd_n}=x$ and n is even, let

$$\alpha_1 = A_0 + z_0P_{yd_2} + A_2 + P_{xd_2}P_{yd_4} + A_4 + \cdots + P_{xd_{2i-2}}P_{yd_{2i}} + A_{2i} + \cdots + P_{xd_{n-2}}P_{yd_n} + A_n,$$

$$\alpha_2 = yP_{yd_1} + A_1 + P_{xd_1}P_{yd_3} + A_3 + \cdots + P_{xd_{2i-1}}P_{yd_{2i+1}} + A_{2i+1} + \cdots + A_{n-1} + P_{xd_{n-1}}x.$$

In this case we see that $\alpha_1+\alpha_2$ is an arc of M from z to y which contains the point x.

If for every integer n, $P_{xd_n}\neq x$, there is a point q which is the sequential limit point of the sequence $P_{xd_1}, P_{xd_2}, P_{xd_3}, \cdots$. We shall show that $q=x$. If not, then there is a component d_q of $M-xy$ satisfying the conditions (a) and (b) with respect to the point q. There is a smallest integer j such that P_{xd_j} is a point of $qP_{yd_q}>$. Then d_{j+1} fails to satisfy condition (b), since P_{xd_j} lies between P_{xd_q} and P_{yd_q} on the arc xy but the arc $yP_{xd_{j+1}}$ does not contain P_{xd_q}. Hence $q=x$. In this case let

$$\alpha_1 = A_0 + z_0P_{yd_2} + \sum_{i=1}^{\infty}(A_{2i} + P_{xd_{2i}}P_{yd_{2i+2}}),$$

$$\alpha_2 = yP_{yd_1} + \sum_{i=1}^{\infty}(A_{2i-1} + P_{xd_{2i-1}}P_{yd_{2i+1}}).$$

Since x is the sequential limit point of the sequences $[P_{xd_{2i}}]$ and $[P_{xd_{2i-1}}]$ and the diameters of the arcs A_i approach zero as i increases, the sets α_1 and α_2 are arcs of M with end points z and x and y and x respectively. It is evident that the arcs α_1 and α_2 have only the point x in common, and thus $\alpha_1+\alpha_2$ is an arc of M with end points z and y and containing the point x.

In any case then, the point x lies on some arc of M whose end points y and z belong to K. But this is contrary to the assumption that $x \subset \overline{M(K)} - M(K)$, which proves our theorem.

THEOREM 3. *If K is any subset of the continuous curve M and α is an arc of M whose end points belong to $\overline{M(K)}$, the arc-curve $M(K)$ contains every point of α except possibly the end points.*

Suppose there exists an arc α of M whose end points belong to $\overline{M(K)}$ such that $<\alpha>$ contains a point z that does not belong to $M(K)$. Since the end points of α belong to $\overline{M(K)}$, the set K contains at least two points. Since z belongs to $<\alpha>$, it is not an end point of M.* Hence either (a) z is a cut point of M or (b) z lies on some simple closed curve of M.† Suppose z is a cut point of M. As $M(K)$ is connected and does not contain z, there is a component H of $M-z$ which contains $M(K)$. As $<\alpha>$ contains z, the point z is not an end point of the continuous curve $H+z$; and as H is connected, z is not a cut point of $H+z$. Hence z lies on a simple closed curve of $H+z$. Then in either case z lies on a simple closed curve J of M, and if z is a cut point of M, the set $J-z$ lies in the component of $M-z$ containing $M(K)$.

By a theorem due to G. T. Whyburn‡ there is a maximal cyclic curve C of M containing the simple closed curve J. If C contains two points x and y of K, let J_{xz} be a simple closed curve of C containing x and z. If the point y lies on J_{xz}, there is an arc of J_{xz} with end points x and y and containing z. Suppose y does not belong to J_{xz}. Now as x is not a cut point of C,§ there exists an arc yvz of $C-x$ with end points y and z. In the order from y to z let t be the first point of J_{xz} on the arc yvz. If $t \neq z$, then the subarc yt of yvz plus that arc of J_{xz} from t to x that contains z is an arc of C from x to y containing z; while if $t=z$, either arc of J_{xz} from x to z may be used together with the subarc yt of yvz. Then if C contains two points x and y of K, there is an arc of C with end points x and y and containing z; and thus z belongs to $M(K)$. Now suppose C contains just one point x of K. Then some component D_y of $M-C$ contains a point y of K. The component D_y has just one limit

---

* Cf. W. L. Ayres, *Concerning continuous curves and correspondences*, Annals of Mathematics, (2), vol. 28 (1927), pp. 396–418, Theorem 4, and G. T. Whyburn, *Concerning continua in the plane*, these Transactions, vol. 29 (1927), pp. 369–400, Theorem 12.

† W. L. Ayres, *Concerning continuous curves and correspondences*, loc. cit., Theorem 3, and G. T. Whyburn, loc. cit., Theorem 22.

‡ *Cyclicly connected continuous curves*, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 31–38, Theorem 3. A continuous curve M is said to be *cyclicly connected* if every two points of M lie together on some simple closed curve of M. A cyclicly connected continuous curve C, which is a subset of a continuous curve M, is said to be a *maximal cyclic curve* of M if M contains no cyclicly connected continuous curve of which C is a proper subset.

§ No point of a cyclicly connected continuous curve is a cut point of the curve. See G. T. Whyburn, *Cyclicly connected continuous curves*, loc. cit.

point w in C.* We may assume that $w \neq x$, for if $w = x$, then, with the Lemma, we could find another component of $M - C$ containing a point of K and whose limit point in C is not x. Since $M(K)$ is a connected subset of $M - z$, the points w and z are distinct. There exists an arc α_1 with end points w and y and lying in D_y except for the point w. As above there exists an arc α_2 of C with end points w and x and containing the point z. The set $\alpha_1 + \alpha_2$ is an arc of M containing z and whose end points x and y belong to K. Hence z belongs to $M(K)$. Finally we suppose that C contains no point of K. Let $[D]$ be the set of all components of $M - C$ that contain points of K, and let $[P]$ be the set of all points of C such that each point P is the limit point of one of the components of $[D]$. If $[P]$ consists of just one point P, then P is a cut point of M, and let

$$M - P = M_1 + M_2,$$

where M_1 and M_2 are non-vacuous mutually separated sets and M_1 is that component of $M - P$ containing $C - P$. Evidently $z \neq P$, for $M(K)$ and $J - P$ lie in different components of $M - P$. Then $\overline{M(K)} \subset M_2 + P$ and, by the Lemma, z cannot belong to the arc α. Thus $[P]$ contains at least two points u_1 and u_2. Let $D_i (i = 1, 2)$ be a component of $[D]$ with u_i as its limit point. The set D_i contains a point v_i of K. There exists an arc β_i whose end points are u_i and v_i and which lies in D_i except for u_i. There exists an arc β_3 of C with end points u_1 and u_2 and containing z. The set $\beta_1 + \beta_2 + \beta_3$ is an arc of M containing z whose end points v_1 and v_2 belong to K. Hence z belongs to $M(K)$. Thus the assumption that z does not belong to $M(K)$ leads to a contradiction in every case.

In Theorem 3 we see that every interior point of an arc of M, whose end points belong to $\overline{M(K)}$, belongs to $M(K)$. We may ask if the end points must belong to $M(K)$ also. The following simple example shows that this is not the case. Let M be the interval whose end points are (0, 0) and (1, 0). Let K be the segment, i.e. the interval minus its end points, from (0, 0) to (1, 0). The points (0, 0) and (1, 0) are points of $\overline{M(K)} - M(K)$, and every interior point of the arc of M joining them is a point of $M(K)$.

THEOREM 4. *If M is a continuous curve and K is a subset of M, the set $\overline{M(K)}$ is identical with the arc-curve $M(\overline{K})$.*

By Theorem 1, $M(K) \subset M(\overline{K})$; and by Theorem 2,

$$\overline{M(K)} - M(K) \subset \overline{K} - K.$$

* G. T. Whyburn, *Cyclicly connected continuous curves*, loc. cit., Theorem 2.

Hence $\overline{M(K)} = M(K) + (\overline{M(K)} - M(K)) \subset M(\overline{K}) + (\overline{K} - K) = M(\overline{K})$. Since $K \subset M(K)$, $\overline{K} \subset \overline{M(K)}$. By Theorem 1, $M(\overline{K}) \subset M(\overline{M(K)})$. But by Theorem 3, $M(\overline{M(K)}) = \overline{M(K)}$. As $\overline{M(K)} \subset M(\overline{K}) \subset M(\overline{M(K)}) = M(K)$, we have $\overline{M(K)} = M(\overline{K})$.

THEOREM 5. *If K is any subset of the continuous curve M, the set $M(K)$ is arc-wise connected im kleinen.**

Let x be any point of $M(K)$, and ϵ be any positive number. Since M is a continuous curve, there exists a positive number $\delta_{\epsilon x}$ such that any point of M whose distance from x is less than $\delta_{\epsilon x}$ can be joined to x by an arc of M whose diameter is less than ϵ. Let y be any point of $M(K)$ whose distance from x is less than $\delta_{\epsilon x}$. There exists an arc α of M with end points x and y and whose diameter is less than ϵ. By Theorem 3, every point of α belongs to $M(K)$, and thus $M(K)$ is arc-wise connected im kleinen.

From Theorems 4 and 5, we have

THEOREM 6. *If K is any subset of the continuous curve M, the set $\overline{M(K)}$ is connected im kleinen.*

Since every point of a continuous curve which is not an end point is an interior point of some arc belonging to the continuous curve, we have from Theorems 3 and 6 the following result: *If K is any subset of the continuous curve M, every point of the set $\overline{M(K)} - M(K)$ is an end point of the continuous curve $\overline{M(K)}$.*

From Theorems 2 and 5, we have the important result

THEOREM 7. *If K is a closed subset of the continuous curve M, the arc-curve $M(K)$ is a continuous curve.*

THEOREM 8. *If x and y are distinct points of the continuous curve M, neither x nor y is a cut point of the arc-curve $M(x+y)$.*

Every point z of $M(x+y)$ lies on some arc A_z of M whose end points are x and y. The subarc yz of A_z is a subset of $M(x+y)-x$. As each point of $M(x+y)-x$ may be joined to y by an arc which belongs to $M(x+y)-x$, the set $M(x+y)-x$ is connected. Similarly $M(x+y)-y$ is connected. Hence neither x nor y is a cut point of $M(x+y)$.

The following example shows that Theorem 8 does not remain true if the set K contains more than two points. Let M be the interval from (0, 0)

* A point set H is said to be *arc-wise connected im kleinen* if for each point x of H and each positive number ϵ, there exists a positive number $\delta_{\epsilon x}$ such that any point of H whose distance from x is less than $\delta_{\epsilon x}$ can be joined to x *by an arc* of H whose diameter is less than ϵ. For closed sets the properties *connected im kleinen* and *arc-wise connected im kleinen* are equivalent.

to (2, 0). Let the set K consist of the three points (0, 0), (1, 0), and (2, 0). In this case the point (1, 0) is a cut point of the arc-curve $M(K)$.

THEOREM 9. *If K is any subset of the continuous curve M, then* (1) *if P is a cut point of $M(K)$ every component of $M(K)-P$ contains at least one point of K,* (2) *every cut point of $M(K)$ is a cut point of M.*

Part (1) is an application of the Lemma. For Part (2), let P be a cut point of $M(K)$ and x and y points of $M(K)-P$ which do not lie in the same component of $M(K)-P$. If $M-P$ is connected, there is an arc α of $M-P$ with end points x and y.* By Theorem 3, every point of α belongs to $M(K)$. Hence x and y belong to the same component of $M(K)-P$, contrary to hypothesis. Then every cut point of $M(K)$ is also a cut point of M.

THEOREM 10. *If K is a closed subset of the continuous curve M, then* (1) *every end point of $M(K)$ belongs to K,* (2) *if T is a maximal connected subset of $M-M(K)$, the set T has only one limit point in $M(K)$,* (3) *if D is a bounded complementary domain of $M(K)$ such that no point of K is interior to the outer boundary of D†, then the boundary of D is a simple closed curve,* (4) *if D is a complementary domain of $M(K)$ there is a complementary domain D_1 of M which is a subset of D and such that every boundary point of D is also a boundary point of D_1.*

Part (1). If P is a point of $M(K)-K$, there exists an arc of M whose end points are points of K and containing P as an interior point. Then P cannot be an end point of $M(K)$ by Theorem 4 of my paper *Concerning continuous curves and correspondences.*‡

Part (2). Suppose T has two limit points P and Q in $M(K)$. Using a theorem due to R. L. Wilder,§ there exists an arc whose end points are P and Q and which lies entirely in T except for the points P and Q. But this is seen to be impossible from Theorem 3.

Part (3) may be proved using a theorem due to R. L. Moore.‖

Part (4) may be proved without difficulty using Part (2) and the result of my paper *Note on a theorem concerning continuous curves.*¶

---

* Cf. R. L. Moore, loc. cit.

† Cf. R. L. Moore, loc. cit., for definition of outer boundary and theorems concerning it.

‡ Loc. cit.

§ Loc. cit.

‖ *Concerning paths that do not separate a given continuous curve*, Proceedings of the National Academy of Sciences, vol. 12(1926), pp. 745–753, Theorem 6.

¶ Loc. cit.

II

Definition. A subset K of a continuous curve M will be called a *basic set* of the continuous curve if $M(K)=M$. A basic set of a continuous curve M will be said to be *irreducible* if it contains no proper subset which is a basic set of M.

THEOREM 11. *In order that a subset K of a continuous curve M be a basic set of M, it is necessary and sufficient that* (1) *K contain at least two points unless M consists of a single point,* (2) *K contain every end point of M,* (3) *if P is a cut point of M, K contain a point of every component of $M-P$.*

The conditions are necessary. The necessity of the first condition is obvious. The necessity of the second and third conditions follows from Theorem 10, Part (1), and Theorem 9, Part (1), respectively, since we assume $M(K)=M$.

The conditions are sufficient. Let K be a subset of M satisfying the three conditions, and let us suppose that $M-M(K)$ is not vacuous. Let x be any point of $M-M(K)$ and y a point of $M(K)$, and let xy denote an arc of M with end points x and y. Let z denote the first point of $\overline{M(K)}$ on the arc xy in the order x to y.

Suppose x is not a limit point of $M(K)$. Then the points x and z are distinct. If z is a cut point of M, let M_1 be the component of $M-z$ containing x, and M_2 be some other component of $M-z$. By condition (3), M_1 and M_2 contain points e_1 and e_2 of K. There exists an arc α_1, whose end points are x and e_1 and every point of which belongs to M_1, and an arc α_2, whose end points are e_2 and z and such that every point of α_2 except z is a point of M_2. Let p be the first point of the subarc xz of the arc xy on the arc α_1 in the order e_1 to x. Since $\alpha_1 \subset M_1$, $p \neq z$. Then the subarc e_1p of α_1 plus the subarc pz of xy plus the arc α_2 is an arc of M whose end points e_1 and e_2 belong to K. Then every point of the subarc pz of xy belongs to $M(K)$. But every point of the set $pz>$ precedes z on the arc xy in the order x to y, which contradicts our assumption that z is the first point of the set $\overline{M(K)}$ on this arc in this order.

If z is not a cut point of M, let w be a point of K distinct from z. The existence of the point w follows from condition (1). As $M-z$ is connected, there is an arc xw with end points x and w, every point of which belongs to $M-z$. In the order x to w let q_1 be the last point of the subarc xz of xy on the arc xw, and let q_2 be the first point of the set $\overline{M(K)}$ on the subarc q_1w of xw. We have $q_2 \neq q_1 \neq z$, and thus the subarc q_2q_1 of xw plus the subarc q_1z of xy is an arc of M which has only its end points in common with $\overline{M(K)}$. But this is impossible by Theorem 3.

Hence every point x of $M-M(K)$ is a limit point of $M(K)$, and thus of K (Theorem 2). As no point of $M-M(K)$ is an end point of M (Condition (2)), the point x is either a cut point of M or lies on some simple closed curve J of M.* If x is a cut point of M, let M_1 and M_2 be distinct components of $M-x$, and let e_1 and e_2 be points of K which belong to M_1 and M_2 respectively. Every arc of M with end points e_1 and e_2 contains x. Hence x belongs to $M(K)$, contrary to assumption. If x lies on a simple closed curve J of M, let y and z be points of J distinct from x and from each other, and let yxz denote the arc of J from y to z which contains the point x. The points y and z belong to $\overline{M(K)}$, since we have shown above that $\overline{M(K)}=M$. Then the point x belongs to $M(K)$, since yxz is an arc of M whose end points belong to $\overline{M(K)}$. Therefore the conditions are sufficient.

THEOREM 12. *In order that a basic set K of a continuous curve M be irreducible, it is necessary and sufficient that there should not exist an arc of M which contains three distinct points of K.*

The condition is necessary. Suppose K is an irreducible basic set of M, and that M contains an arc α which contains three distinct points x, y, z of K. One of these points, and we will suppose x is the one, lies between the other two on the arc α. We shall prove that $K-x$ is a basic set of M and thus show that K is reducible, by showing that $K-x$ satisfies the three conditions of Theorem 11.

The set $K-x$ satisfies condition (1) since it contains the two points y and z. The set K contains every end point of M since it is a basic set of M, and x is not an end point of M since it is interior to an arc of M, namely the arc α. Hence $K-x$ contains every end point of M and condition (2) is satisfied. If $K-x$ does not satisfy condition (3), there is a cut point P of M and a component M_1 of $M-P$ which contains no point of $K-x$. Since K satisfies condition (3), M_1 contains the single point x of K. Then, by the Lemma, the point x cannot be interior to the subarc yz of α. Thus $K-x$ satisfies all three conditions of Theorem 11 and is a basic set of M.

The condition is sufficient. Suppose K contains a proper subset H which is a basic set of M. Let x be a point of $K-H$. Since $x \subset M-H=M(H)-H$, there exists an arc β of M, which contains x and whose end points y and z are points of H. Then the arc β of M contains three distinct points x, y and z of K.

THEOREM 13. *In order that a basic set K of a continuous curve M be irreducible it is necessary and sufficient that* (1) *K contain no cut point of M,*

* See my paper, *Concerning continuous curves and correspondences*, loc. cit., Theorem 3. See also G. T. Whyburn, *Concerning continua in the plane*, loc. cit., Theorem 22.

(2) *if C is a maximal cyclic curve of M, the set C contain two points, one point or no point of K according as C contains no cut point, one cut point or more than one cut point of M.*

The conditions are necessary. Let K be an irreducible basic set of the continuous curve M. If K contains a point z which is a cut point of M, let M_1 and M_2 be two distinct components of $M-z$. Since K is a basic set, by Theorem 11 the component M_i $(i=1, 2)$ contains a point x_i of K. Let α denote an arc of M whose end points are x_1 and x_2. Since M_1 and M_2 are components of $M-z$, the arc α contains the point z. Then K is not irreducible by Theorem 12.

If C is a maximal cyclic curve of M and contains no cut point of M, then $C=M$.* Then K consists of just two points, for if K contains three or more points, there exists an arc of M containing three points of K† and, by Theorem 12, the set K is not an irreducible basic set. If C contains just one cut point P of M, the set $C-P$ is a component of $M-P$ and thus contains at least one point of K by Theorem 11. If $C-P$ contains two points x and y of K, there is an arc α_1 belonging to C which contains y and whose end points are P and x. Let M_1 be a component of $M-P$ which is distinct from $C-P$, and let z be a point of K which belongs to M_1 (Theorem 11). There exists an arc α_2 whose end points are P and z and which lies in M_1 except for the point P. Then $\alpha_1+\alpha_2$ is an arc of M which contains three points x, y, z of K. But this is impossible since K is an irreducible basic set (Theorem 12). If C contains two cut points P and Q of M, let M_1 be a component of $M-P$ which does not contain $C-P$, and let M_2 be a component of $M-Q$ which does not contain $C-P$. It may be seen easily that M_1 and M_2 have no common points. Let x and y be points of K which belong to M_1 and M_2 respectively (Theorem 11). Now suppose C contains a point z of K. Since condition (1) is necessary, $P\neq z\neq Q$. There exist arcs α_1 and α_2 with end points x and P and y and Q respectively, and such that α_1 lies in M_1 except for the point P and α_2 lies in M_2 except for the point Q. There exists an arc α_3 with end points P and Q and such that

$$z \subset \alpha_3 \subset C.$$

Then $\alpha_1+\alpha_2+\alpha_3$ is an arc of M containing three points x, y, and z of K. But this is impossible, for K is an irreducible basic set of M.

---

* Cf. G. T. Whyburn, *Cyclicly connected continuous curves*, loc. cit., Theorem 1.

† In the proof of Theorem 3 we showed that if x, y and z are any three points of a cyclicly connected continuous curve C, there is an arc of C with end points x and y and containing z.

The conditions are sufficient. Let K be a basic set of M satisfying the two conditions, and suppose that K is reducible. Then K contains a proper subset H which is a basic set of M. Let x be a point of $K-H$. By Theorem 11, the set H contains every end point of M. Then the point x is not an end point of M. By condition (1), the set K contains no cut point of M; and since $x \subset K$, the point x is not a cut point of M. Hence the point x must lie on some simple closed curve of M, and there is a maximal cyclic curve C containing the point x.* If C contains no cut point of M, every point of H belongs to C. Since H contains at least two points (Theorem 11) and $H \subset K$, the maximal cyclic curve C contains three points of K, contrary to condition (2). If C contains one cut point P of M, the set $C-P$ is a component of $M-P$ and contains at least one point of H (Theorem 11). Then the curve C contains two points of K, contradicting condition (2). If C contains more than one cut point of M, C contains the point x of K, thus contradicting condition (2). Hence the conditions are sufficient.

It is evident that every basic set of a bounded continuous curve contains a subset which is an irreducible basic set. Also there are continuous curves which have more than one irreducible basic set.

---

* Cf. G. T. Whyburn, *Cyclicly connected continuous curves*, loc. cit.

THE UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

A SOLUTION OF THE MATRIC EQUATION $P(X)=A$*

BY

WILLIAM E. ROTH

I. INTRODUCTION

The equation $P(X)=A$, where $P(\lambda)$ is a polynomial in λ with scalar coefficients and A is a square matrix of order n, has received the attention of various writers. Perhaps the first to deal with the solution of such an equation of degree greater than the first was Cayley in his *Memoir on the theory of matrices.*† He there gave a solution for the equation $L=M^{1/2}$, where M was a known matrix of order two or of order three. The theory expounded in the remarkable memoirs of Cayley was further developed by Sylvester;‡ he gave a general solution of the equation $X^p=A$, but he did not give the deductions that led him to his results, nor did he discuss the conditions under which his solution applies. He asserted that the solutions of $X^p=A$ are p^μ in number, where μ is the number of distinct roots of the characteristic equation of A. The statement is correct for the kind of solutions he gave, namely, those expressible as polynomials in the given matrix. He recognized the existence of solutions not so expressible in case $X^p=I$, where I is the identical matrix, and later treated this particular case separately.§ In an article‖ which appeared in 1883, he called attention to the relationship of quaternions to matrices of the second order and gave a definition of the four units of quaternions in terms of matrices; and from then on he took up the discussion of quadratic equations in quaternions.¶

The work of Sylvester advanced the subject considerably; but the increased interest in mathematical foundations and in logical rigor led to new

---

* Presented to the Society, September 9, 1927; received by the editors December 19, 1927.

† See Philosophical Transactions of the Royal Society of London, vol. 148 (1858), pp. 17–37; or Collected Mathematical Papers, vol. II, pp. 475–496.

‡ Sylvester, *Sur les puissances et les racines de substitutions linéaires*, Comptes Rendus, vol. 94 (1882), pp. 55–59; or Mathematical Papers, vol. III, pp. 562–4.

§ Sylvester, *Sur les racines des matrices unitaires*, Comptes Rendus, vol. 94 (1882), pp. 396–9; or Mathematical Papers, vol. III, pp. 565–7.

‖ Sylvester, *On the involution and evolution of quaternions*, Philosophical Magazine, vol. 16 (1883), pp. 394–396; or Mathematical Papers, vol. IV, pp. 112–114.

¶ Sylvester, *Sur la solution explicite de l'équation quadratique de Hamilton en quaternions ou en matrices du second ordre*, Comptes Rendus, vol. 99 (1884), pp. 555–8, 621–631; or Mathematical Papers, vol. IV, pp. 188–198.

treatments of some of his problems; the principal contributions to the algebra of matrices from the modern point of view were made by Frobenius. His solution of the binomial equation,* $X^2=A$, where A is a non-singular square matrix, has found its way into modern textbooks.† Dickson‡ gives practically the same solution extended to any degree in X. Frobenius stated that his solution may readily be extended to apply when A is a singular matrix, but how this can be done is not clear from his discussion and apparently has never been accomplished by his method. The solutions of Frobenius are p^μ in number and are expressible as polynomials in the given matrix; in both these respects his results agree with those of Sylvester.

The introduction of the Weierstrass§ elementary divisors opened a new mode of attack upon the problems we are considering here. This was employed by Kreis‖ in his general solution of the equation $P(X)=A$ defined above, for which he gave solutions that are expressible as polynomials in the given matrix; his results are expressed in terms of the Weierstrass elementary divisors and associated normal forms. Later he¶ treated the binomial equation $X^p=A$ separately and gave a criterion for the existence of solutions when A is non-singular or singular. A course similar to that of Kreis was followed by Cecioni,** who solved the equation $X^p=A$, but not the more general equation, $P(X)=A$. Cecioni calls the solutions formed by Frobenius "soluzioni singolari" and says that these and only these can be expressed as linear aggregates of powers of the given matrix; he gives a criterion by means of which one should be able to determine the existence of such solutions. In this, however, he seems to have been led into an error as will be pointed out later. He further considers the possible solutions in a field F which contains the elements of A. The paper of Cecioni, like those of Kreis, is very difficult to read because of the difficulties involved by the use of elementary divisors.

---

* Frobenius, *Über die cogredienten Transformation der bilinearen Formen*, Sitzungsbericht der Königlichen Preussischen Akademie der Wissenschaften, 1896.

† See Muth, *Theorie und Anwendung der Elementarteiler*, Leipzig, Teubner, 1899. Bôcher, *Introduction to Higher Algebra*, New York, Macmillan, 1907.

‡ Dickson, *Modern Algebraic Theories*, Chicago, Sanborn, 1926.

§ Weierstrass, *Zur Theorie der bilinearen und quadratischen Formen*, Monatsberichte der Königlichen Preussischen Akademie der Wissenschaften, 1868, pp. 310–338.

‖ Kreis, *Contribution à la Théorie des Systèmes linéaires*, Zürich Thesis, 1906.

¶ Kreis, *Auflösung der Gleichung* $X^m=A$, Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich, 53te Jahrgang, 1908, pp. 366–376.

** Cecioni, *Sopra alcune operazioni algebriche sulle matrici*, Annali della Reale Scuola Normale Superiore di Pisa, vol. 11 (1909), pp. 1–40.

The present paper will give such solutions of the equation $P(X)=A$ as are expressible as polynomials in A, and a criterion for the existence of such solutions will be established. The method here to be developed has an advantage over that of Frobenius, in that it applies as well in certain cases where A is singular and is not restricted to the binomial equation. It is based upon the given equation and the polynomial $\psi(\lambda)$, for which $\psi(A)=0$, which may or may not be the equation of lowest degree satisfied by A. The use of infinite expansions such as were used by Frobenius and others is entirely avoided, thus removing the doubt which must arise when that method is used, in as much as a series in a given matrix may not be convergent when it is convergent for scalar quantitites. However, the method does not give all the solutions that may occur when there exists an equation $\psi(A)=0$ of degree lower than the order of A. This fact was pointed out by Sylvester and becomes evident if we regard the equation $P(X)=kI$; the solutions of this equation expressible as polynomials in I must necessarily have the form $X=\alpha I$, where α is a root of the equation $P(\lambda)=k$, whereas other solutions are known to exist according to the results of Kreis and Cecioni cited above.

Frobenius* gave two theorems that are closely related to Theorem II of the present paper but they cannot be used to prove the existence of a solution for the equation $P(X)=A$. At any rate Frobenius did not refer to them when he solved the equation $X^2=A$, even though the theorems were published previous to the appearance of his solution of this equation.

II. Preliminary theorems

THEOREM I. *If $\psi(\lambda)$ is a polynomial of degree $m>1$ in λ, and the distinct roots of $\psi(\lambda)=0$ are α_j, $j=1, 2, \cdots, s$; if $P(\lambda)$ is a polynomial in λ of degree $p>1$, whose leading coefficient is unity and whose constant term is zero; and if the equation $P(\lambda)-\alpha_j=0$ has at least one simple root for each α_j, $j=1, 2, \cdots, s$, which is a multiple root of $\psi(\lambda)=0$; then polynomials $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, of degree m in λ exist such that*

$$\prod_{i=1}^{p} \phi_i(\lambda) \equiv \psi(P(\lambda)),$$

and such that at least one, $\phi_k(\lambda)$, $1 \leqq k \leqq p$, has no quadratic factor in λ in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$.

---

* Frobenius, *Über lineare Substitutionen und bilineare Formen*, Crelle's Journal, vol. 84 (1878), pp. 1–63.

Suppose that $\psi(\lambda)$ is given by the identity

$$\psi(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \alpha_j)^{\epsilon_j}, \tag{1}$$

where α_j, $j=1, 2, \cdots, s$, are the distinct roots of $\psi(\lambda)=0$ and where $\sum_{j=1}^{s} \epsilon_j = m$. We have, by hypothesis,

$$P(\lambda) \equiv \lambda^p + h_1\lambda^{p-1} + h_2\lambda^{p-2} + \cdots + h_{p-1}\lambda\,; \tag{2}$$

and we assume further that $P(\lambda)-\alpha_j, j=1, 2, \cdots, s$, is given by

$$P(\lambda) - \alpha_j \equiv \prod_{i=1}^{p}(\lambda - \beta_{ij}) \qquad (j = 1,2, \cdots, s)\,; \tag{3}$$

then $-h_1, h_2, \cdots, (-1)^{p-1}h_{p-1}, (-1)^{p-1}\alpha_j$ are the elementary symmetric functions of the roots β_{ij}, $i=1, 2, \cdots, p$, of the equations $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$. From (1) and (3), it follows that

$$\begin{aligned}\psi(P(\lambda)) &\equiv \prod_{j=1}^{s}(P(\lambda) - \alpha_j)^{\epsilon_j},\\ &\equiv \prod_{j=1}^{s}\Big[\prod_{i=1}^{p}(\lambda - \beta_{ij})\Big]^{\epsilon_j}.\end{aligned}$$

The factors of the right member of this identity may be rearranged in a number of ways so as to give

$$\psi(P) \equiv \prod_{i=1}^{p}\Big[\prod_{j=1}^{s}(\lambda - \beta_{ij})^{\epsilon_j}\Big], \tag{4}$$

where here and in the following the abbreviated notation, $\psi(P)$, will be used to denote $\psi(P(\lambda))$ regarded as a polynomial in λ unless the contrary is specifically stated, and the corresponding notation will be used with the same significance for other polynomials in $P(\lambda)$. Now let

$$\phi_i(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \beta_{ij})^{\epsilon_j} \qquad (i = 1,2, \cdots, p)\,; \tag{5}$$

then $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, will be polynomials of degree $\sum_{j=1}^{s} \epsilon_j = m$ in λ; and by (4),

$$\psi(P) \equiv \prod_{i=1}^{p}\phi_i(\lambda). \tag{6}$$

It still remains to be shown that at least one $\phi_k(\lambda)$, $1 \leqq k \leqq p$, has no factor of the second degree in λ in common with any polynomial $P(\lambda)-\alpha_j, j=1, 2,$

$\cdots, s$. To this end we write the roots of each equation $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$, in a separate column thus:

$$\begin{array}{cccc} P(\lambda)-\alpha_1=0, & P(\lambda)-\alpha_2=0, & \cdots, & P(\lambda)-\alpha_s=0; \\ \beta_{11}, & \beta_{12}, & \cdots, & \beta_{1s}; \\ \beta_{21}, & \beta_{22}, & \cdots, & \beta_{2s}; \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{p1}, & \beta_{p2}, & \cdots, & \beta_{ps}. \end{array}$$

Then the β's occurring in any one column are distinct from those occurring in any other column; for suppose $\beta_{rh}=\beta_{tk}$, $h\neq k$; then by (3), $P(\beta_{rh})=\alpha_h$, and $P(\beta_{tk})=\alpha_k$, but since $\beta_{rh}=\beta_{tk}$, it follows that $\alpha_h=\alpha_k$, $h\neq k$. This is impossible because α's with different subscripts are distinct. Then no β of any one column of the above table is repeated in any other column. On the other hand, β's in the same column are not necessarily distinct from each other. The definition of $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, given in (5) shows that $\lambda-\beta_{ij}$, $j=1, 2, \cdots, s$, has the same exponent, ϵ_j, in $\phi_i(\lambda)$ that the corresponding factor, $\lambda-\alpha_j$, has in $\psi(\lambda)$; consequently it may be said that a certain polynomial $\phi_k(\lambda)$, $1\leqq k\leqq p$, is formed by taking for its factors one β from each column of the table above; each factor $\lambda-\beta_{kj}$ so taken is given the same exponent ϵ_j in $\phi_k(\lambda)$ that $\lambda-\alpha_j$ has in $\psi(\lambda)$. Thus there is a one-to-one correspondence between the factors $(\lambda-\beta_{kj})^{\epsilon_j}$ of $\phi_k(\lambda)$ and $(\lambda-\alpha_j)^{\epsilon_j}$ of $\psi(\lambda)$. The only way in which any $\phi_k(\lambda)$ can have a factor of the second degree in λ in common with some $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, is that $\lambda-\beta_{kj}$ be a multiple factor of $P(\lambda)-\alpha_j$, where $\lambda-\alpha_j$ is a multiple factor of $\psi(\lambda)$. But by hypothesis we know that $P(\lambda)-\alpha_j$ has at least one simple factor, when $\lambda-\alpha_j$ is a multiple factor of $\psi(\lambda)$. Consequently it is possible to construct at least one polynomial $\phi_k(\lambda)$ which will have no factor of the second degree in λ in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$. The remaining polynomials $\phi_i(\lambda)$, $i=1, 2, \cdots, k-1, k+1, \cdots, p$, must then be formed in one of $(p-1)^s$ ways, not necessarily all distinct, from the $p-1$ elements remaining in each column of the above table; together with $\phi_k(\lambda)$, they will satisfy identity (6).

In general, we can say the polynomial $\phi_k(\lambda)$, having the above properties, may be formed in $\prod_{j=1}^{s}\mu_j$ distinct ways, where s is the number of distinct roots of $\psi(\lambda)=0$, and where μ_j is the number of distinct roots of $P(\lambda)-\alpha_j=0$ when α_j is a simple root of $\psi(\lambda)=0$ and the number of simple roots of $P(\lambda)-\alpha_j=0$ when α_j is a multiple root of $\psi(\lambda)=0$. If some polynomial $P(\lambda)-\alpha_j$ has only multiple factors when α_j is a multiple root of $\psi(\lambda)=0$, then the

corresponding $\mu_j=0$ and no polynomial $\phi_k(\lambda)$ can be formed satisfying the conditions above.

THEOREM II. *Under the conditions of Theorem* I *and with the polynomial,* $\phi_k(\lambda)$, *whose existence was there established, there exist polynomials* $H_k(\lambda)$, $Z_k(\lambda)$, *and* $T_k(\lambda)$, *the latter of degree less than* m *in* λ, *such that*

$$H_k(\lambda)\phi_k(\lambda) \equiv \lambda - T_k(P(\lambda)),$$

and that

$$Z_k(\lambda)\psi(\lambda) \equiv \lambda - P(T_k(\lambda)).$$

We have, as under the preceding theorem,

$$\psi(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \alpha_j)^{\epsilon_j}, \tag{1}$$

where the α_j, $j=1, 2, \cdots, s$, are distinct;

$$P(\beta_{kj}) = \alpha_j \qquad (j = 1,2, \cdots, s); \tag{7}$$

and

$$\phi_k(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \beta_{kj})^{\epsilon_j}. \tag{5}$$

Furthermore, $\phi_k(\lambda)$ has no quadratic factor in common with any polynomial $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, and because of (6) we can write the identity

$$\psi(P) \equiv \phi_k(\lambda)Q(\lambda), \tag{8}$$

where $Q(\lambda)$ is a polynomial in λ.

We shall show that every polynomial $\psi_1(\lambda)$, not identically zero, and such that

$$\psi_1(P) \equiv \phi_k(\lambda)Q_1(\lambda), \tag{9}$$

is divisible by $\psi(\lambda)$, and consequently there exists no polynomial $\psi_1(\lambda)$ of degree lower than m in λ which can satisfy an identity of this kind. Let $\psi_1(\lambda)$ be any polynomial that satisfies the identity (9), and substitute β_{kj}, $j=1, 2, \cdots, s$, for λ in this identity; then because of (5) and (7),

$$\psi_1(\alpha_j) = 0 \qquad (j = 1,2, \cdots, s). \tag{10}$$

That is, the distinct factors of $\psi(\lambda)$ must also be factors of $\psi_1(\lambda)$. If we can now show that the multiplicity of any factor $(\lambda-\alpha_\tau)$, $1\leqq\tau\leqq s$, of $\psi_1(\lambda)$ is ϵ_τ, then according to (1) $\psi_1(\lambda)$ must be divisible by $\psi(\lambda)$. In case all factors of $\psi(\lambda)$ are simple, this assertion is already proved. If $(\lambda-\alpha_\tau)^{\epsilon_\tau}$, $\epsilon_\tau>1$, is a factor of $\psi(\lambda)$, then $(\lambda-\beta_{k\tau})^{\epsilon_\tau}$ is a factor of $\phi_k(\lambda)$; we shall now show

that $\psi_1(\lambda)$ must have the factor $(\lambda-\alpha_\tau)^{\epsilon_\tau}$ in common with $\psi(\lambda)$. For this purpose we differentiate the members of (9) with respect to λ, and we find

$$\psi_1'(P)P'(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}'.$$

Substitute $\beta_{k\tau}$ for λ in this identity; since $\epsilon_\tau>1$, the right member is zero and we have

$$\psi_1'(\alpha_\tau)P'(\beta_{k\tau}) = 0.$$

Now $(\lambda-\beta_{k\tau})^{\epsilon_\tau}$, $\epsilon_\tau>1$, is a factor of $\phi_k(\lambda)$ and since $P(\lambda)-\alpha_\tau$ has the factor $\lambda-\beta_{k\tau}$, $P'(\lambda)$ cannot have this factor, for the polynomials $P(\lambda)-\alpha_\tau$ and $\phi_k(\lambda)$ have no quadratic factor in λ in common. Therefore

$$P'(\beta_{k\tau}) \neq 0, \tag{11}$$

and we must have

$$\psi_1'(\alpha_\tau) = 0. \tag{12}$$

Thus $(\lambda-\alpha_\tau)^2$ is a factor of $\psi_1(\lambda)$. If now $\epsilon_\tau>2$, we take the second derivatives with respect to λ of the members of (9), and thus obtain the identity

$$\psi_1''(P)P'^2(\lambda) + \psi_1'(P)P''(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}''.$$

Substitute $\beta_{k\tau}$ for λ in this identity; then the right member is zero and because of (11) and (12)

$$\psi_1''(\alpha_\tau) = 0.$$

This equation together with (10) and (12) permits us to conclude that $(\lambda-\alpha_\tau)^3$ is a factor of $\psi_1(\lambda)$, if $\epsilon_\tau>2$. To show that this procedure may be continued step by step to justify the conclusion that $\psi_1(\lambda)$ has the factor $(\lambda-\alpha_\tau)^{\epsilon_\tau}$, $1\leqq\tau\leqq s$, we assume that rth derivatives of the members of (9) with respect to λ satisfy the identity

$$\begin{aligned}\psi_1^{(r)}(P)P'^r(\lambda) + \psi_1^{(r-1)}(P)R_r(\lambda) + \psi_1^{(r-2)}(P)S_r(\lambda) + \cdots \\ + \psi_1''(P)U_r(\lambda) + \psi_1'(P)P^{(r)}(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}^{(r)},\end{aligned} \tag{13}$$

in which the leading term of the left member is $\psi_1^{(r)}(P)\,P'^r(\lambda)$ and the remaining terms are in $\psi_1^{(r-1)}(P)$, $\psi_1^{(r-2)}(P)$, $\cdots$, $\psi_1(P)$ multiplied by polynomials in λ. This formula holds for $r=1$ and for $r=2$; to show that it holds in general we differentiate its members with respect to λ. This gives us

$$\begin{aligned}\psi_1^{(r+1)}(P)P'^{r+1}(\lambda) &+ \psi_1^{(r)}(P)[rP'^{r-1}(\lambda)P''(\lambda) + P'(\lambda)R_r(\lambda)] \\ &+ \psi_1^{(r-1)}(P)[R_r'(\lambda) + P'(\lambda)S_r(\lambda)] + \cdots \\ &+ \psi_1''(P)[U_r'(\lambda) + P'(\lambda)P^{(r)}(\lambda)] + \psi_1'(P)P^{(r)}(\lambda) \\ &\equiv \{\phi_k(\lambda)Q_1(\lambda)\}^{(r+1)},\end{aligned}$$

which is again of the general form (13) with r replaced by $r+1$. That formula is therefore valid.

Now we assume we have shown by successive steps that

$$\psi_1(\alpha_\tau) = \psi_1'(\alpha_\tau) = \psi_1''(\alpha_\tau) = \cdots = \psi_1^{(r-1)}(\alpha_\tau) = 0,$$

where $r \leqq \epsilon_\tau - 1$; then letting $\lambda = \beta_{k\tau}$ in (13), the right member will be zero for $(\lambda - \beta_{k\tau})^{\epsilon_\tau}$ is a factor of $\phi_k(\lambda)$ and because of the relations just written and because of (11), we must likewise have

$$\psi_1^{(r)}(\alpha_\tau) = 0.$$

Then $(\lambda - \alpha_\tau)^{r+1}$ must be a factor of $\psi_1(\lambda)$, if $(\lambda - \alpha_\tau)^r$ is and if $(\lambda - \beta_{k\tau})^{\epsilon_\tau}$, $\epsilon_\tau \geqq r+1$, is a factor of $\phi_k(\lambda)$. We are therefore permitted to conclude that $\psi_1(\lambda)$ has the factors $(\lambda - \alpha_\tau)^{\epsilon_\tau}$, $1 \leqq \tau \leqq s$, and because $\alpha_1, \alpha_2, \cdots, \alpha_s$ are distinct, that $\psi_1(\lambda)$ is divisible by $\prod_{j=1}^{s} (\lambda - \alpha_j)^{\epsilon_j}$, that is, by $\psi(\lambda)$. Furthermore, no polynomial $\psi_1(\lambda)$ of lower degree than m, the degree of $\psi(\lambda)$, exists which can satisfy an identity of the form given by (9); and any $\psi_1(\lambda)$ of degree m which satisfies identity (9) must be of the form $\psi_1(\lambda) \equiv c\psi(\lambda)$, where c is an arbitrary constant not zero, and is therefore uniquely determined save for a constant factor. We are now ready to proceed with the proof of the present theorem.

Let $\beta_{k\tau}$ be any root of $\phi_k(\lambda) = 0$; dividing $[P(\lambda) - \alpha_\tau]^\sigma$ by $\phi_k(\lambda)$, we find for every positive integer σ

$$[P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) Q_\sigma(\lambda) + (\lambda - \beta_{k\tau}) R_\sigma(\lambda), \tag{14}$$

where $Q_\sigma(\lambda)$ and $R_\sigma(\lambda)$ are polynomials in λ, and the degree of $R_\sigma(\lambda)$ in λ will not exceed $m-2$; $\lambda - \beta_{k\tau}$ is a factor of the remainder because it is a factor of both $P(\lambda) - \alpha_\tau$ and $\phi_k(\lambda)$. For $\sigma = m-1$, the left side of (14) is a polynomial of degree $p(m-1)$. Since we supposed $p > 1$ and $m > 1$, it follows that $p(m-1) \geqq m$, the degree of $\phi_k(\lambda)$, so that

$$Q_{m-1}(\lambda) \not\equiv 0.$$

On the basis of (14) we form the sum

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda) + (\lambda - \beta_{k\tau}) \sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda), \tag{15}$$

where t_σ, $\sigma = 1, 2, \cdots, m-1$, are arbitrary constants. If it were possible to choose $t_1, t_2, \cdots, t_{m-1}$ not all zero so that

$$\sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda) \equiv 0, \tag{16}$$

we would have for these values of $t_1, t_2, \cdots, t_{m-1}$

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda),$$

so that $\sum_{\sigma=1}^{m-1} t_\sigma(\lambda-\alpha_\tau)^\sigma$ would be a polynomial of degree less than m in λ satisfying the identity (9). Since no such polynomial can satisfy that condition, (16) is impossible, and $R_\sigma(\lambda)$, $\sigma=1, 2, \cdots, m-1$, are linearly independent. This fact permits us to define the constants $t_1, t_2, \cdots, t_{m-1}$ so that

$$\sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda) \equiv 1. \tag{17}$$

If we suppose that

$$R_\sigma(\lambda) \equiv r_{1\sigma}\lambda^{m-2} + r_{2\sigma}\lambda^{m-3} + \cdots + r_{m-1,\sigma},$$

then t_σ, $\sigma=1, 2, \cdots, m-1$, must satisfy the $m-1$ non-homogeneou equations

$$\sum_{\sigma=1}^{m-1} r_{i\sigma} t_\sigma = 0 \qquad (i = 1,2, \cdots, m-2),$$

$$\sum_{\sigma=1}^{m-1} r_{m-1,\sigma} t_\sigma = 1.$$

The determinant of the coefficients of this system of equations is

$$\begin{vmatrix} r_{11} & r_{12} & \cdots & r_{1\,m-1} \\ r_{21} & r_{22} & \cdots & r_{2\,m-1} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m-1\,1} & r_{m-1\,2} & \cdots & r_{m-1\,m-1} \end{vmatrix},$$

and is not zero, for $R_1(\lambda)$, $R_2(\lambda)$, $\cdots$, $R_{m-1}(\lambda)$ whose coefficients form the elements of the first, second, $\cdots$, and last column of this determinant are linearly independent. Consequently the constants $t_1, t_2, \cdots, t_{m-1}$ may be uniquely determined so that (17) is satisfied, and (15) may be written as

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda) + \lambda - \beta_{k\tau}.$$

If we let

$$T_k(\lambda) \equiv \beta_{k\tau} + \sum_{\sigma=1}^{m-1} t_\sigma(\lambda - \alpha_\tau)^\sigma, \tag{18}$$

and

$$H_k(\lambda) \equiv -\sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda),$$

this identity becomes

$$H_k(\lambda)\phi_k(\lambda) \equiv \lambda - T_k(P(\lambda)), \tag{19}$$

where $T_k(\lambda)$ is of degree lower than m in λ.

Finally, by (3),

$$P(\lambda) - \alpha_r \equiv \prod_{i=1}^{p}(\lambda - \beta_{ir}),$$

and if here we replace λ in the right member by $H_k(\lambda)\phi_k(\lambda)+T_k(P)$ according to (19), we obtain the identity

$$P(\lambda) - \alpha_r \equiv P(T_k(P)) - \alpha_r + K_k(\lambda)\phi_k(\lambda),$$

or

$$K_k(\lambda)\phi_k(\lambda) \equiv P(\lambda) - P(T_k(P)),$$

where $K_k(\lambda)\phi_k(\lambda)$ is the aggregate of all terms of the product that contain the factor $\phi_k(\lambda)$ explicitly. This identity is of the form (9), where $\lambda - P(T_k(\lambda))$ replaces $\psi_1(\lambda)$; $\lambda - P(T_k(\lambda))$ is consequently divisible by $\psi(\lambda)$, so that we may write

$$Z_k(\lambda)\psi(\lambda) \equiv \lambda - P(T_k(\lambda)), \tag{20}$$

where $Z_k(\lambda)$ is a polynomial in λ. This completes the proof of the present theorem.

III. Solution of the equation $P(X)=A$

THEOREM III. *If $\psi(\lambda)$ is a polynomial of degree $m>1$ in λ, and the distinct roots of $\psi(\lambda)=0$ are α_j, $j=1, 2, \cdots, s$; if $P(\lambda)$ is a polynomial of degree $p>1$ in λ whose leading coefficient is unity and whose constant term is zero; if the equation $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$, has at least one simple root for every α_j which is a multiple root of $\psi(\lambda)=0$; and if*

$$\psi(A) = 0,$$

where A is a square matrix of order n; then there exists at least one matrix X also of order n such that

$$P(X) = A,$$

*and such that X is expressible as a polynomial in A with scalar coefficients.**

* The theorem holds as well for $m=1$ and for $p=1$, but in either case the results would be trivial and do not merit separate treatment.

Under the same hypotheses as those here stated, we have proved in Theorem II that a polynomial $T_k(\lambda)$ exists such that

$$\lambda - P(T_k(\lambda)) \equiv \psi(\lambda)Z_k(\lambda), \tag{20}$$

where $Z_k(\lambda)$ is a polynomial in λ. According to the present theorem we have $\psi(A)=0$, and since powers of a single matrix and their products with scalars obey the commutative and distributive laws of ordinary algebra, it is evident that if λ above be replaced by A and λ^0 by I, we have

$$A - P(T_k(A)) = 0,$$

or

$$P(T_k(A)) = A.$$

Therefore the equation

$$P(X) = A$$

has a solution which is given by

$$X_k = T_k(A), \tag{21}$$

where $T_k(\lambda)$ is a polynomial in λ with scalar coefficients. This completes the proof of our theorem.*

The existence of X is not only proved in this theorem but its form in terms of A is explicitly given by (21). Indeed the solution X_k of the equation $P(X)=A$ corresponding to the polynomial $\phi_k(\lambda)$, which has no factor of the second degree in common with any polynomial $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, is distinct from that corresponding to any other such polynomial of the $\prod_{j=1}^{s}\mu_j$ that may have this property. (See page 583.) Before we prove the uniqueness of the solutions, we shall show that

$$\phi_k(X_k) = 0.$$

According to (5) and (19),

$$\phi_k(\lambda) \equiv \prod_{j=1}^{s}[T_k(P) - \beta_{kj} + H_k(\lambda)\phi_k(\lambda)],$$

$$\equiv \phi_k(T_k(P)) + L_k(\lambda)H_k(\lambda)\phi_k(\lambda),$$

* If $\psi(\lambda)$ is the polynomial of lowest degree such that $\psi(A)=0$, then the conditions of the present theorem are also necessary for the existence of X expressible as a polynomial in A. The writer's attention was called to this fact by Professor Ingraham. The logic by which we may prove that the conditions are necessary is virtually given by Frobenius, *Über lineare Substitutionen und bilineare Formen*, Crelle's Journal, vol. 84 (1878), p. 13, if we regard his $g(r)$ as the known function, $f(r)$ as the unknown function, and $\psi(r)$ as the polynomial of lowest degree such that $\psi(A)=0$. In our notation these polynomials would be $P(\lambda)$, $T_k(\lambda)$, and $\psi(\lambda)$ respectively. The method leads to the construction of $T_k(\lambda)$ by means of solving for the coefficients of $T_k(\lambda)$ a linear system of equations with non-zero determinant.

where $L_k(\lambda)H_k(\lambda)\phi_k(\lambda)$ is the aggregate of all terms of the product that have $H_k(\lambda)\phi_k(\lambda)$ as a factor. This identity is again of the form (9); in the present case all terms save $\phi_k(T_k(P))$, which corresponds to $\psi_1(P(\lambda))$ of that identity, contain $\phi_k(\lambda)$ explicitly and therefore $\phi_k(T_k(\lambda))$ is divisible by $\psi(\lambda)$. We may therefore write the identity above in the form

$$\phi_k(\lambda) \equiv \psi(P)M_k(P) + L_k(\lambda)H_k(\lambda)\phi_k(\lambda),$$

where $M_k(\lambda)$ is a polynomial in λ. Now by (19) and (21)

$$H_k(X_k)\phi_k(X_k) = X_k - T_k(A) = 0,$$

and since $\psi(A)=0$, then by substituting X_k for λ and I for λ^0 in the identity established above, we have

$$\phi_k(X_k) = 0,$$

for X_k is a solution of the equation $P(X)=A$. Consequently we have shown that $\phi_k(\lambda)$ is a polynomial that is satisfied by the matrix X_k obtained by the above method.

Thus far no restriction upon the polynomial $\psi(\lambda)$, save that $\psi(A)=0$, has been made. However, in order to prove the uniqueness of the solution obtained for each polynomial $\phi_k(\lambda)$ of the permitted class, we assume that $\psi(\lambda)$ of degree m is the polynomial of lowest degree satisfied by A. Suppose further that $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are two distinct polynomials of the $\prod_{j=1}^{s}\mu_j$ that have no quadratic factor in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, and that determine the same solution X for the equation $P(X)=A$. Then we would have

$$\phi_k(X) = 0, \quad \phi_l(X) = 0;$$

two equations of degree m satisfied by X. Consequently it will be possible to determine constants c_1 and c_2 so that

$$F(\lambda) \equiv c_1\phi_k(\lambda) + c_2\phi_l(\lambda),$$

where $F(\lambda)$ will be a polynomial not identically zero and of degree $m'<m$, since $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are distinct. But substituting X for λ and I for λ^0 we have $F(X)=0$. Then X satisfies a polynomial of degree m'. Suppose

$$F(\lambda) \equiv \prod_{j=1}^{m'}(\lambda - \gamma_i)$$

and that

$$P(\gamma_j) = \delta_j \qquad (j = 1,2, \cdots, m'),$$

where neither the γ_j nor the δ_j are necessarily distinct. Let

$$\Psi(\lambda) \equiv \prod_{j=1}^{m'} (\lambda - \delta_j) ;$$

now according to the definition of δ_j, each polynomial $P(\lambda)-\delta_j$ will have $\lambda-\gamma_j$ as a factor and the product of the m' polynomials $P(\lambda)-\delta_j$, $j=1, \cdots, m'$, will be divisible by $F(\lambda)$, i.e.,

$$\Psi(P) \equiv \prod_{j=1}^{m'} (P(\lambda) - \delta_j),$$

$$\equiv F(\lambda)N(\lambda),$$

where $N(\lambda)$ is a polynomial in λ. Now making the usual substitution X for λ we have

$$\Psi(A) = F(X)N(X) = 0,$$

since $P(X)=A$ and $F(X)=0$. But $\Psi(\lambda)$ is a polynomial of degree m', whereas the polynomial $\psi(\lambda)$ of degree m was assumed as that of lowest degree which vanished for A, so that the above equation is impossible. Hence the solutions X_k and X_l corresponding to distinct polynomials $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are distinct.

The number of distinct solutions of the equation $P(X)=A$, expressible as polynomials in A, is therefore given by $\prod_{j=1}^{s} \mu_j$, where s is the number of distinct roots of the equation $\psi(\lambda)=0$, where $\psi(\lambda)$ is the polynomial of lowest degree for which $\psi(A)=0$, and where μ_j is the number of distinct roots of $P(\lambda)-\alpha_j=0$, when α_j is a simple root of $\psi(\lambda)=0$, and the number of simple roots of $P(\lambda)-\alpha_j=0$, when α_j is a multiple root of $\psi(\lambda)=0$.

Though the solution obtained by the method developed above is not restricted by the condition that $\psi(\lambda)$ is the polynomial of lowest degree for which $\psi(A)=0$, there is clearly no advantage in employing a polynomial of higher degree since such a course would only increase the labor involved in solving a particular example and may not give all the solutions that are possible. This last statement becomes evident if we recall that it is entirely possible in certain cases for α_r to be a simple root of $\psi(\lambda)=0$, where $\psi(\lambda)$ is of lower degree than n, the order of A, and where $\psi(A)=0$, but for the characteristic equation of A to have α_r as a multiple root. If in such a case $P(\lambda)-\alpha_r=0$ have multiple roots the characteristic equation would permit fewer solutions by the above method than would be possible on the basis of $\psi(\lambda)=0$.

The equation $X^p = A$ has p^s solutions when A is non-singular, for then all roots of the equation $\lambda^p = \alpha_j$ are distinct and $\mu_j = p$, $j = 1, 2, \cdots, s$. On the other hand, when A is singular and the equation $\psi(\lambda) = 0$, where $\psi(\lambda)$ is such that $\psi(A) = 0$, has a simple root $\alpha_1 = 0$, then the equation $X^p = A$ has p^{s-1} solutions of the kind here studied; for in this case the equation $\lambda^p = \alpha_1$ has only the root $\lambda = 0$, hence $\mu_1 = 1$, whereas the other $s-1$ equations, $\lambda^p - \alpha_j = 0$, $j = 2, 3, \cdots, s$, will each have p distinct solutions; if, however, $\alpha_1 = 0$ is a multiple root of $\psi(\lambda) = 0$, where $\psi(A) = 0$ is the equation of lowest degree satisfied by A, then since $\lambda^p - \alpha_1 = 0$ has no simple roots, we see that according to the definition of μ_j above, $\mu_1 = 0$, and thus our method would in this case give no solutions of the equation $X^p = A$.

The conclusion here reached that the equation $X^p = A$, $|A| = 0$, always has solutions if a polynomial $\psi(\lambda)$, such that $\psi(A) = 0$, exists which does not have λ^2 as a factor, contradicts a statement made by Cecioni.* This contradiction will be exhibited explicitly by means of a numerical example. (See Example 2 below.)

It may be shown that for a given matrix A, singular or non-singular, the coefficients $h_1, h_2 \cdots, h_{p-2}$ of the polynomial

$$P(\lambda) \equiv \lambda^p + h_1 \lambda^{p-1} + h_2 \lambda^{p-2} + \cdots + h_{p-2} \lambda^2 + h_{p-1} \lambda$$

may be taken entirely arbitrarily and h_{p-1} may be chosen in an infinity of ways so that the equation $P(X) = A$ will have solutions expressible as polynomials in the given matrix A.

The solution developed above for the equation $P(X) = A$ permits the following conclusion. If the elements of A belong to the algebraic field F which contains the coefficients of $P(\lambda)$ and if $\psi(\lambda)$ is the characteristic function or the polynomial of lowest degree such that $\psi(A) = 0$, then the solutions of the equation $P(X) = A$ that are expressible as polynomials in A belong to the field formed by the adjunction of the roots of $\psi(P(\lambda)) = 0$ to the field F. This conclusion follows from the fact that the coefficients of $\psi(\lambda)$ belong to the field that contains the elements of A, and that the coefficients of $\phi_k(\lambda)$ belong to the field containing the roots of $\psi(P(\lambda)) = 0$.

IV. Examples

Example 1. Given

$$P(X) \equiv X^2 - 3X = A,$$

* Cecioni, loc. cit., p. 85.

where

$$A = \begin{bmatrix} -3 & 3 & 6 \\ -\frac{7}{3} & 5 & 2 \\ -\frac{4}{3} & 4 & 0 \end{bmatrix};$$

X must be found satisfying this equation.

The polynomial $\psi(\lambda)$, such that $\psi(A)=0$, is in this case

$$\psi(\lambda) \equiv \lambda^3 - 2\lambda^2 - 8\lambda.$$

The roots of $\psi(\lambda)=0$ are 0, -2, and 4. The table giving the roots of $P(\lambda)-\alpha_j=0$, $j=1, 2, 3$, is

$P(\lambda)=0$,	$P(\lambda)+2=0$,	$P(\lambda)-4=0$,
0,	1,	-1,
3,	2,	4.

$\phi_k(\lambda)$ may be formed in the following 2^3 ways:

$$\phi_1(\lambda) \equiv \lambda(\lambda-1)(\lambda+1), \quad \phi_2(\lambda) \equiv (\lambda-3)(\lambda-2)(\lambda-4),$$
$$\phi_1'(\lambda) \equiv \lambda(\lambda-1)(\lambda-4), \quad \phi_2'(\lambda) \equiv (\lambda-3)(\lambda-2)(\lambda+1),$$
$$\phi_1''(\lambda) \equiv \lambda(\lambda-2)(\lambda+4), \quad \phi_2''(\lambda) \equiv (\lambda-3)(\lambda-1)(\lambda+1),$$
$$\phi_1'''(\lambda) \equiv \lambda(\lambda-2)(\lambda+1), \quad \phi_2'''(\lambda) \equiv (\lambda-3)(\lambda-1)(\lambda-4).$$

Each of the polynomials above is such that it has no quadratic factor in common with any polynomial $P(\lambda)$, $P(\lambda)+2$, $P(\lambda)-4$, and consequently for each there exists a unique matrix X such that $P(X)=A$ and $\phi_i^{(r)}(X)=0$, $i=1, 2$, $r=0, 1, 2, 3$. The solutions may be obtained directly from the pairs of equations $P(X)=A$ and $\phi_i^{(r)}(X)=0$, and this process would be entirely legitimate after the results above. We will, however, for the present develop the solution that satisfies $\phi_2(\lambda)$ to illustrate the theory as developed above.

In this case $\phi_2(\lambda)\equiv\lambda^3-9\lambda^2+26\lambda-24$, and according to (16), where in the present case $P(\lambda)-\alpha\equiv\lambda^2-3\lambda+2$, we have

$$P(\lambda)+2 \equiv (\lambda-2)(\lambda-1),$$
$$[P(\lambda)+2]^2 \equiv \phi_2(\lambda)[\lambda+3] + (\lambda-2)(14\lambda-38).$$

Then

$$\sum_{\sigma=1}^{2} t_\sigma R_\sigma(\lambda) \equiv t_1(\lambda-1) + t_2(14\lambda-38) \equiv 1,$$

and consequently

$$t_1 = 14/24, \quad t_2 = -1/24.$$

Then

$$T_2(\lambda) \equiv 2 + \frac{14}{24}(\lambda + 2) - \frac{1}{24}(\lambda + 2)^2,$$

$$\equiv -\frac{1}{24}(\lambda^2 - 10\lambda - 72),$$

and we consequently must have

$$X_2 = -\frac{1}{24}(A^2 - 10A - 72I)$$

as a solution of $P(X) = A$, according to the above theory. In the same way we get the solutions corresponding to each of the polynomials above. These are here tabulated in order:

$$X_1 = \frac{1}{24}(A^2 - 10A), \qquad X_2 = -\frac{1}{24}(A^2 - 10A - 72I),$$

$$X_1' = \frac{1}{4}A^2, \qquad X_2' = -\frac{1}{4}(A^2 - 12I),$$

$$X_1'' = \frac{1}{3}(A^2 - A), \qquad X_2'' = -\frac{1}{3}(A^2 - A - 9I),$$

$$X_1''' = \frac{1}{8}(A^2 - 6A), \qquad X_2''' = -\frac{1}{8}(A^2 - 6A - 24I),$$

where $X_i^{(r)}$ is the solution whose characteristic function in each case is $\phi_i^{(r)}(\lambda)$, $i = 1, 2$, and $r = 0, 1, 2, 3$.

EXAMPLE 2. We propose to find solutions of the equation $X^2 = A$, where

$$A = \begin{pmatrix} 1 & -8 & 0 & 8 \\ 1 & 1 & 0 & -1 \\ 4 & 0 & 4 & -4 \\ 1 & -8 & 0 & 8 \end{pmatrix}$$

is a singular matrix.

The characteristic function of A is a polynomial $\psi(\lambda)$ of lowest degree such that $\psi(A) = 0$. The roots of $\psi(\lambda) \equiv \lambda^4 - 14\lambda^3 + 49\lambda^2 - 36\lambda = 0$ are $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = 4$, $\alpha_4 = 9$. The roots of $\lambda^2 - \alpha_j = 0$, $j = 1, 2, 3, 4$, are

$$\begin{array}{cccc} \lambda^2=0, & \lambda^2-1=0, & \lambda^2-4=0, & \lambda^2-9=0, \\ 0, & 1, & 2, & 3, \\ 0, & -1, & -2, & -3; \end{array}$$

whence the following polynomials ϕ are formed:

$$\begin{aligned} \phi_1(\lambda) &\equiv \lambda(\lambda-1)(\lambda-2)(\lambda-3), & \phi_2(\lambda) &\equiv \lambda(\lambda+1)(\lambda+2)(\lambda+3), \\ \phi_1'(\lambda) &\equiv \lambda(\lambda-1)(\lambda-2)(\lambda+3), & \phi_2'(\lambda) &\equiv \lambda(\lambda+1)(\lambda+2)(\lambda-3), \\ \phi_1''(\lambda) &\equiv \lambda(\lambda-1)(\lambda+2)(\lambda-3), & \phi_2''(\lambda) &\equiv \lambda(\lambda+1)(\lambda-2)(\lambda+3), \\ \phi_1'''(\lambda) &\equiv \lambda(\lambda+1)(\lambda-2)(\lambda-3), & \phi_2'''(\lambda) &\equiv \lambda(\lambda-1)(\lambda+2)(\lambda+3). \end{aligned}$$

Each one of these polynomials is such that it has not a factor of the second degree in common with any $\lambda^2-\alpha_j, j=1, 2, 3, 4$. Here we have, according to our method,

$$\begin{aligned} \lambda^2 &\equiv && \lambda^2, \\ \lambda^4 &\equiv \phi_1(\lambda) && +\lambda(6\lambda^2-11\lambda+6), \\ \lambda^6 &\equiv \phi_1(\lambda)[\lambda^2+6\lambda+25] && +\lambda(90\lambda^2-239\lambda+150). \end{aligned}$$

Then

$$\sum_{\sigma=1}^{3} t_\sigma R_\sigma(\lambda) \equiv t_1\lambda + t_2(6\lambda^2-11\lambda+6) + t_3(90\lambda^2-239\lambda+150) \equiv 1.$$

and

$$t_1 = 74/60,\ t_2 = -15/60,\ t_3 = 1/60.$$

Then

$$T_1(\lambda) \equiv \frac{1}{60}(\lambda^3-15\lambda^2+74\lambda),$$

or

$$X_1 = \frac{1}{60}(A^3-15A^2+74A).$$

X_1 is a solution of $\phi_1(X)=0$ and $X^2=A$, and its negative is a solution of $\phi_2(X)=0$ and $X^2=A$. In this manner we get as the remaining solutions:

$$X_1' = \mp\frac{1}{6}(A^2-7A),$$

$$X_1'' = \mp\frac{1}{12}(A^3-11A^2+22A),$$

$$X_1''' = \mp\frac{1}{30}(2A^3-25A^2+53A),$$

where the upper sign is that of the solution obtained for $\phi_1^{(r)}(X)=0$ and $X^2=A$ and the lower of that for $\phi_2^{(r)}(X)=0$ and $X^2=A$. These solutions are, respectively,

$$X_1 = -X_2 = \begin{pmatrix} 1 & -2 & 0 & 2 \\ 1 & 1 & 0 & -1 \\ 2 & 0 & 2 & -2 \\ 1 & -2 & 0 & 2 \end{pmatrix}, \quad X_1' = -X_2' = \begin{pmatrix} 1 & 4 & 0 & -4 \\ 1 & 1 & 0 & -1 \\ 2 & 0 & 2 & -2 \\ 1 & 4 & 0 & -4 \end{pmatrix},$$

$$X_1'' = -X_2'' = \begin{pmatrix} 1 & -2 & 0 & 2 \\ 1 & 1 & 0 & -1 \\ -2 & 0 & -2 & 2 \\ 1 & -2 & 0 & 2 \end{pmatrix}, \quad X_1''' = -X_2''' = \begin{pmatrix} -1 & -4 & 0 & 4 \\ -1 & -1 & 0 & 1 \\ 2 & 0 & 2 & -2 \\ -1 & -4 & 0 & 4 \end{pmatrix}.$$

Thus the conclusion reached in the preceding section, that the equation $X^p=A$ may have solutions expressible as polynomials in A even when $|A|=0$, is verified by a particular example when $p=2$.

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

CONCERNING THE CUT POINTS OF CONTINUA*

BY

GORDON T. WHYBURN

In this paper propositions will be established concerning the collection of all the cut points of a given plane† continuum. It will be shown that there does not exist an uncountable collection of mutually exclusive subcontinua of a given continuum M each of which contains at least one cut point of M. With the aid of this fundamental theorem it is shown, among other things, that the set of all the cut points of a continuum M which are irregular points of M is necessarily countable and, indeed, that all save a countable number of the cut points of any continuum M are points of Menger order two of M.

1. Preliminary notions. Intervals and segments of a continuum

Definitions. The points, or point sets, A and B of a continuum M are said to be separated in M by the point X of M provided that $M-X$ is the sum of two mutually separated point sets $M_a(X)$ and $M_b(X)$ containing A and B respectively. The notation $K(A, B)$ will be used to denote the set of all points of M which separate A and B in M.

Let A and B be any two points of a continuum M and X and Y any two distinct points of $K(A, B)$. Then X will be said to *precede* Y on M in the order from A to B provided that X belongs to the set $M_a(Y)$; X will be said to follow Y provided that X belongs to the set $M_b(Y)$. It is easily seen that if X precedes Y, then Y follows X. If X precedes Y, then by the *interval of M from X to Y* (or from Y to X), or the *interval XY* (or YX) *of M*, will be meant the set of points $[M_b(X)+X]\cdot[M_a(Y)+Y]$, and by the *segment* of M from X to Y will be meant the set of points $M_b(X)\cdot M_a(Y)$. If X follows Y, then since Y precedes X, by the interval XY of M is meant the set of points $[M_b(Y)+Y]\cdot[M_a(X)+X]$, and the segment XY (or YX) of M means the set of points $M_a(X)\cdot M_b(Y)$. The notation $I(X, Y)$ for the interval of M from X to Y and $S(X, Y)$ for the segment of M from X to Y will be used in this paper. Clearly $S(X, Y)=I(X, Y)-(X+Y)$.

The following two propositions can be proved with little difficulty from the above definitions.

* Presented to the Society, September 9, 1927; received by the editors in September, 1927.

† Quite recently I have established some very general results on the cuttings of continua, as a consequence of which it follows that all the theorems in this paper hold true in a euclidean space of n dimensions, and nearly all of them hold in any separable metric space which is locally compact. See my forthcoming paper *Concerning collections of cuttings of connected point sets.*

THEOREM 1. *Let A and B be any two points of a continuum M, and let X and Y be any two distinct points of $K(A, B)$, where X precedes Y. Then M is the sum of three continua $M_a(X)+X$, $I(X, Y)$, and $M_b(Y)+Y$, where $[M_a(X)+X]\cdot I(X, Y)=X$, $[M_b(Y)+Y]\cdot I(X, Y)=Y$, and $[M_a(X)+X]\cdot[M_b(Y)+Y]=0$.*

THEOREM 2. *Using the same notation as in Theorem 1, suppose that each of the sets $M_a(X)$, $M_b(X)$, $M_a(Y)$, and $M_b(Y)$ is connected. Then if $S(X, Y)$ contains a point of $K(A, B)$, $S(X, Y)$ is connected.*

2. THE PRINCIPAL RESULTS

THEOREM 3. *If H is any uncountable subset of the set G of all the cut points of a continuum M, then there exist two points A and B of H which are separated in M by uncountably many of the points of H.**

There exists a countable subset D of M such that every point of M is a limit point of D. Let F be the collection of all possible pairs of points of D. Then F is countable. For each pair of points u, v of F, let $K_h(u, v)$ denote the set of all points of H that separate u and v in M. Then $H=\sum_F K_h(u, v)$. For let X be a point of H. Then since X is a cut point of M, $M-X=M_1+M_2$, where M_1 and M_2 are mutually separated point sets. Since every point of M is a limit point of D, it follows that M_1 contains a point U of D and M_2 contains a point V of D. Then X separates U and V in M, and hence it belongs to $K_h(U, V)$. Therefore $H=\sum_F K_h(u, v)$; and since F is countable and H is uncountable, it follows that there exists at least one pair P, Q of F such that $K_h(P, Q)$ is uncountable.

Let E denote the set of points $K_h(P, Q)$. Then since E is uncountable, it contains a point A which is a point of condensation of E, i.e., every vicinity of A contains uncountably many points of E. Now for at least one of the sets $M_p(A)$ and $M_q(A)$, say for $M_q(A)$, it is true that A is a point of condensation of the subset of E which belongs to $M_q(A)$. Let B be a point of E belonging to $M_q(A)$. Then by Theorem 1 it readily follows that the segment AB of M must contain an uncountable subset E_0 of E. Let X be any point of E_0. Then since X belongs to $S(A, B)$, it follows by Theorem 1 that neither of the continua $M_p(A)+A$ and $M_q(B)+B$ contains the point X. And as these continua contain the points P and Q respectively, therefore they are subsets of $M_p(X)$ and $M_q(X)$ respectively. Hence A and B are separated in M by each point X of E_0, and as E_0 is an uncountable subset of H, the truth of Theorem 3 is therefore established.

* The referee has kindly called my attention to the fact that Theorems 3 and 4 of this paper are valid in any number of dimensions, since my proofs of these theorems make no use of the dimensionality of the space in which the point sets lie.

THEOREM 4. *Let G denote the set of all the cut points of a continuum M. Then for all, save possibly a countable number, of points X of G it is true that $M-X$ is the sum of two mutually separated and connected point sets.**

Suppose, on the contrary, that there exists an uncountable subset D of G such that if X is any point of D, then $M-X=M_1(X)+M_2(X)$, where $M_1(X)$ and $M_2(X)$ are mutually separated point sets and at least one of them is not connected. Then by Theorem 3 there exist two points A and B of D which are separated in M by each point of an uncountable subset H of the set D. For each point X of H, $M-X=H_a(X)+H_b(X)+H_0(X)$, where $H_a(X)$, $H_b(X)$, and $H_0(X)$ are mutually separated point sets, and $H_a(X)$ and $H_b(X)$ contain A and B respectively.

Now for each point X of H, the set H, and indeed the set $K(A, B)$, is a subset of the set of points $H_a(X)+H_b(X)+X$, because this set of points is connected and contains both A and B and therefore must contain every point of M which separates A and B in M. It follows that if X and Y are distinct points of H, the set of points $H_0(Y)$ belongs to the set $H_a(X)+H_b(X)$, and the set $H_0(X)$ belongs to the set $H_a(Y)+H_b(Y)$; and therefore the sets $H_0(X)$ and $H_0(Y)$ are mutually exclusive. By the Zermelo postulate there exists a set of points E such that for each point X of H there is exactly one point P_x of E which belongs to the set of points $H_0(X)$, and for each point P_x of E there is exactly one point X of H such that the set $H_0(X)$ contains the point P_x. Since H is uncountable, E must be uncountable. Hence there exists a point P_z of E which is a limit point of E. But there exists a point Z in H such that P_z belongs to the set $H_0(Z)$, and $E-P_z$ belongs to the set of points $H_a(Z)+H_b(Z)$. Clearly this is impossible, because $H_a(Z)$, $H_b(Z)$, and $H_0(Z)$ are mutually separated point sets. Thus the supposition that Theorem 4 is not true leads to a contradiction.

Theorem 4 can also be proved very readily with the aid of a statement given, without proof, by C. Zarankiewicz in his paper *Sur les points de division dans les ensembles connexes.*†

THEOREM 5. *Let M be any continuum whatever. Then there does not exist an uncountable collection of mutually exclusive subcontinua‡ of M each of which contains at least one cut point of M.*

---

* Since this paper was submitted to the editors, a result somewhat more general than Theorem 4 has been published by Kuratowski and Zarankiewicz; cf. *A theorem on connected point sets*, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 571–575.

† Cf. Fundamenta Mathematicae, vol. 9, pp. 124–171, footnote to p. 137.

‡ In this paper the hypothesis that a set is a continuum is understood to imply that the set contains more than one point.

Before proceeding with the proof of Theorem 5, I shall establish two lemmas which will be needed in the proof.

LEMMA 1. *If C is any subcontinuum of a continuum M, then not more than a countable number of the bounded maximal connected subsets of $M-C$ can contain cut points of M.*

Suppose, on the contrary, that there exists an uncountable collection G of bounded maximal connected subsets of $M-C$ each of which contains at least one cut point of M. From each element g of G select exactly one cut point P_g of M, and let K denote the set of all points $[P_g]$ thus selected. For each point P_g of K, $M-P_g=M_c(P_g)+M_0(P_g)$, where $M_c(P_g)$ and $M_0(P_g)$ are mutually separated sets and $M_c(P_g)$ contains the continuum C. Let H denote the sum of all the point sets of the collection G. Then since each element of G has* at least one limit point in C, it follows that for each point P_g of K, the set of points $H-g$ belongs to the set $M_c(P_g)$. By the Zermelo postulate, there exists a set of points E such that for each point P_g of K there is exactly one point Q_g of E which belongs to the set of points $M_0(P_g)$, and for each point Q_g of E there is exactly one point P_g of K such that the set of points $M_0(P_g)$ contains the point Q_g. It is readily seen that no two of the sets $[M_0(P_g)]$ can have a common point. Therefore, as H is uncountable, both K and E must be uncountable. Hence E contains a point Q_j which is a limit point of E. But clearly this is impossible, because there exists a point P_j in K such that Q_j belongs to the set $M_0(P_j)$ and $E-Q_j$ belongs to the set $M_c(P_j)$, and $M_0(P_j)$ and $M_c(P_j)$ are mutually separated sets by definition. This contradiction proves Lemma 1.

LEMMA 2. *If M is any continuum, C is any simple closed curve, and G_0 denotes the collection of all the maximal connected subsets of M contained in C plus its interior, then not more than a countable number of elements of G_0 can contain cut points of M which lie within C.*

Suppose, on the contrary, that G_0 contains an uncountable subcollection G each element g of which contains at least one cut point P_g of M lying within C.

(1) Not more than a countable number of elements g of G can have the property that $M-P_g$ can be expressed as the sum of two mutually separated sets one of which lies wholly within C. For suppose the contrary is true. Then, clearly, for each such element g of G, the point P_g is a cut point of the continuum $M+C$. But since each such point P_g lies on a bounded maximal

* See my paper *Concerning point sets which can be made connected by the addition of a simple continuous arc*, these Transactions, vol. 29 (1927), footnote to p. 754.

connected subset of $(M+C)-C$, and no two different points $[P_g]$ can lie on the same maximal connected subset of $(M+C)-C$, this is contradictory to Lemma 1.

(2) Let G_1 denote the collection of all those elements g of G such that however $M-P_g$ be expressed as the sum of two mutually separated point sets $M_1(P_g)$ and $M_2(P_g)$, each of these sets must contain at least one point of C. It follows by (1) that G_1 must be uncountable. It is clear that if g is any element of G_1 and we let $g_1(P)=g\cdot M_1(P_g)$ and $g_2(P)=g\cdot M_2(P_g)$, then $g-P_g=g_1(P)+g_2(P)$, and $g_1(P)$ and $g_2(P)$ are mutually separated point sets each of which contains at least one point on C. Now C is not a subset of M, for if it were, then G_0 would have only one element. Let the maximal segments of $C-M\cdot C$ be denoted by $S_1, S_2, S_3, \cdots$. Let g be any element of G_1. Denote the end points of S_1 by A and B, and let AOB be the arc of C from A to B which does not contain S_1. On AOB, in the order from A to B, let X_g be the first point belonging to g. Then X_g belongs either to $g_1(P)$ or to $g_2(P)$, say to $g_1(P)$. Then on the arc X_gB of AOB, in the order from X_g to B, let Y_g be the first point belonging to $g_2(P)$. It is easy to see that the segment X_gY_g of the arc AOB contains at least one segment S_i of the sequence $S_1, S_2, S_3, \cdots$. Let A_i and B_i denote the end points of S_i, where A_i precedes B_i on AOB in the order from A to B. Then if T_1 and T_2 denote the arcs AA_i and B_iB respectively of AOB, $g_1(P)$ has at least one point on one of these arcs and $g_2(P)$ has at least one point on the other. Thus it is seen that for each element g of G_1 there exist two segments of the sequence $S_1, S_2, S_3, \cdots$ such that $g_1(P)$ has a point on one of the complementary arcs of C of these two segments and $g_2(P)$ has a point on the other. Then since the number of segments $S_1, S_2, S_3, \cdots$ is countable, and G_1 is uncountable, there exist two segments S_k and S_j of the sequence $S_1, S_2, S_3, \cdots$ such that if T and L denote the complementary arcs of C, then for uncountably many elements g of G_1 it is true that $g_1(P)$ has at least one point on one of the arcs T and L, and $g_2(P)$ has at least one point on the other. Let G_2 denote the collection of all those elements of G_1 which have this property.

(3) There are not more than a countable number of elements g of G_2 such that $T+L$ contains as many as three points of g. For suppose the contrary is true. Then there exists an uncountable subcollection E of G_2 such that each element e of E has at least two points on one of the arcs T and L, say on T, and at least one point on L. Let A and B denote the end points of T, where A is an end point of S_k, and for each element e of E, let A_e and B_e respectively denote the first and last points on T in the order from A to B which belong to e. Then for no two elements e and f of E can the segments

A_eB_e and A_fB_f of T have a common point. For suppose they do. Then either the segment A_eB_e contains a point of f or the segment A_fB_f contains a point of e. The two cases are alike, so let us suppose that the segment A_eB_e contains a point X of f. Now L contains at least one point Y of f. There exists an arc XZY from X to Y which lies, except for the points X and Y, wholly without C. It is readily seen that the segments S_k and S_j belong to different complementary domains R_k and R_j respectively of the continuum $f+XZY$. Since A_e can be joined to A and B_e to B by arcs lying, except for their end points, in the exterior of C and containing no point whatever of XZY, it follows that A_e must belong to R_k and B_e to R_j. Then since e is connected, it must have at least one point in common with $f+XZY$. But as the segment XZY lies without C, e can have no point in common with it. Hence e must have a point in common with f. This is contradictory to the fact that e and f are different maximal connected subsets of M lying in C plus its interior. Therefore no two segments A_eB_e and A_fB_f can have a common point. But since E is uncountable, then the number of these segments must be uncountable. Clearly this is impossible. This contradiction proves the statement made in the first sentence of this paragraph.

(4) It follows from (3) that there exists an uncountable subcollection G_3 of elements of G_2 such that if e is any element of G_3, then $e_1(P)$ has exactly one point X_e on T and $e_2(P)$ has exactly one point Y_e on L. But since G_3 is uncountable, it contains an element g which contains the sequential limiting set H of some infinite sequence V of elements of G_3. Then H is connected and contains at least one point on each of the arcs T and L. But since g has in common with $T+L$ only the points X_g and Y_g, H must contain both X_g and Y_g. Now some infinite subsequence V_1 of V has the property that all of its elements belong to one of the sets $M_1(P_g)$ and $M_2(P_g)$, say to $M_1(P_g)$. But H is also the limiting set of the sequence V_1, and the point Y_g belongs to the set $M_2(P_g)$. Hence $M_2(P_g)$ contains a limit point of $M_1(P_g)$, contrary to the fact that $M_1(P_g)$ and $M_2(P_g)$ are mutually separated point sets. Thus the supposition that Lemma 2 is not true leads to a contradiction.

Proof of Theorem 5. Suppose, contrary to Theorem 5, that there exists an uncountable collection G of mutually exclusive subcontinua of M each of which contains at least one cut point of M. From each element g of G select exactly one cut point P_g of M, and let N denote the set of all points $[P_g]$ thus selected. Since N is uncountable, there exists a square S which encloses uncountably many points of N. Let N_1 be the set of all those points of N which are within S. There exists a positive number d and an uncountable subset N_2 of N_1 such that if G_2 denotes the collection of all those elements

of G which contain a point of N_2, then each element of G_2 is of diameter greater than $4d$. Let us divide S plus its interior into a finite collection E of squares plus their interiors by lines parallel and perpendicular to the bases of S in such a way that the diameter of each square of E is less than d. Then either some side T of one of the squares of E contains uncountably many points of N_2, or else some square of E encloses uncountably many of these points. In either case it is readily seen that there exists a rectangle R_1 of diameter less than $2d$ which encloses uncountably many of the points of N_2. There exists a rectangle R_2 concentric with and within R_1 which also encloses an uncountable subset N_3 of N_2.

Let F denote the collection of all the maximal connected subsets of M contained in R_2 plus its interior. Since N_3 is uncountable and each point of N_3 belongs to some element of F, and since, by Lemma 2, not more than a countable number of elements of F can contain points of N_3, it follows that some element K of F must contain an uncountable subset N_4 of N_3. It follows by a theorem of R. L. Moore's* that uncountably many points of N_4 must be cut points also of the continuum K. Let N_5 be the set of all those points of N_4 which are cut points of K. By Theorem 3, there exist two points A and B of N_5 which are separated in K and also in M by uncountably many points of N_5. Let N_6 be the set of all those points of N_5 which separate A and B in K and also in M, and let G_6 be the collection of all those elements of G_2 which contain a point of N_6. Since each element of G_6 is of diameter greater than $4d$, and R_1 is of diameter less than $2d$, each element of G_6 must contain a point of R_1 and also a point of R_2. For each element g of G_6, the maximal connected subset H_{0g} of g which contains the point P_g and lies within R_1 has at least one limit point A_{P_g} which belongs to R_1. Let H_{P_g} denote the continuum obtained by adding to H_{0g} all of its limit points. Now since A and B belong to N_5, the elements g_a and g_b of G_2 which contain A and B respectively contain subcontinua H_a and H_b respectively which contain A and B respectively and at least one point U and V respectively on R_1 but which lie wholly in R_1 plus its interior. It follows by Theorem 4 that N_6 contains an uncountable subset N_7 such that if X is any point of N_7, then $M-X=M_a(X)+M_b(X)$, where $M_a(X)$ and $M_b(X)$ are mutually separated and connected point sets containing A and B respectively, and also $K-X=K_a(X)+K_b(X)$, where $K_a(X)$ and $K_b(X)$ are connected sets belonging to $M_a(X)$ and $M_b(X)$ respectively. Let E_0 denote the collection of all the continua $[H_x]$ selected above which correspond to points X of N_7. At least one of the arcs of R_1 from U to V must contain uncountably many points

* R. L. Moore, *Concerning the cut points of continuous curves and of other closed and connected point sets*, Proceedings of the National Academy of Sciences, vol. 9 (1923), pp. 101–106.

A_x which correspond to points X of N_7 and belong to continua H_x of the collection E_0; let UOV denote one which does, and let N_8 be the set of all those points of N_7 such that the corresponding point A_x belongs to the arc UOV.

Now if X and Y are two points of N_8 such that X precedes Y on K in the order from A to B (see §1), I shall show that the point A_x precedes the point A_y on the arc UOV in the order from U to V. Suppose, on the contrary, that A_y precedes A_x on UOV in the order from U to V. Then U and A_x separate the points V and A_y on R_1, and if UWA_x is an arc from U to A_x which lies, except for its end points, in the exterior of R_1, then it is easy to see that the points A_y and V lie in different complementary domains of the continuum $UWA_x+H_a+K_a(X)+H_x$. Then since the continuum $H_b+K_b(Y)+H_y$ contains both A_y and V, therefore it must contain at least one point in common with $UWA_x+H_a+K_a(X)+H_x$. This is impossible, for no one of the sets H_b, $K_b(Y)$, and H_y can have a point in common with any one of the sets UWA_x, H_a, $K_a(X)$, and H_x. Thus the supposition that A_x does not precede A_y on the arc UOV in the order from U to V leads to a contradiction.

Now since N_8 is uncountable, the set of corresponding points $[A_x]$ is uncountable. Accordingly there exists one of these points A_z which is a limit point of a subset S_1 of the remaining ones belonging to the arc UA_z of UOV and also of a subset S_2 of the remaining ones belonging to the arc A_zV of UOV. Now the point A_z belongs either to $M_a(Z)$ or to $M_b(Z)$, say to $M_a(Z)$. But this is impossible, since A_z is a limit point of S_2, and if A_x is any point of S_2, it was shown above that the point X follows the point Z on M in the order from A to B (for the order on M and on K is the same), i.e., X belongs to $M_b(Z)$; and thus A_z is a limit point of $M_b(Z)$, which is absurd. Likewise we arrive at a contradiction if we suppose that A_z belongs to $M_b(Z)$. Thus the supposition that Theorem 5 is not true leads to a contradiction and the theorem is therefore established.

THEOREM 6. *If E is any subset of the set of all the cut points of a continuum M, then there are not more than a countable number of points X of E such that X belongs to some subcontinuum of M which contains no other point of E.*

Suppose, on the contrary, that there exists an uncountable subset D of E such that each point X of D belongs to some subcontinuum N_x of M containing no other point of E. Then by Theorems 3 and 4 it follows that there exist two points A and B of M and an uncountable subset F of D such that if X is any point of F, X separates A and B in M and furthermore each of the sets $M_a(X)$ and $M_b(X)$ is connected.

Now for not more than a countable number of points X of F is it true that there exists some other point Y of F such that the segment XY of M contains no point whatever of F. For let Q be the set of all points X of F such that there exists a point Y of F such that $S(X, Y)$ contains no point of F, and suppose, contrary to the statement just made, that Q is uncountable. It follows that there exists an uncountable collection of the segments $[S(X, Y)]$ no two of which are identical and no one of which contains a point of F. Let $S(X_1, Y_1)$ and $S(X_2, Y_2)$ be any two of these segments, where X_1 precedes Y_1 and X_2 precedes Y_2 on M in the order from A to B. Then $S(X_1, Y_1)$ and $S(X_2, Y_2)$ are mutually exclusive. For suppose they have a point P in common. One of these segments is not a subset of the other one, for they are not identical. Suppose $S(X_2, Y_2)$ is not a subset of $S(X_1, Y_1)$. Let K denote the set of points common to $S(X_1, Y_1)$ and $S(X_2, Y_2)$. Now not both of the points X_1 and Y_1 can be limit points of K; for if they were, they would both belong to $I(X_2, Y_2)$ and at least one of them would have to belong to $S(X_2, Y_2)$, contrary to the fact that $S(X_2, Y_2)$ contains no point of F. The two cases are alike, so let us suppose X_1 is not a limit point of K. Then it is easy to see that K is closed except possibly for the point Y_1. Since K belongs to $S(X_1, Y_1)$, it must belong to $M_a(Y_1)$. But $M_a(Y_1)$ is connected; and since $M_a(Y_1)-K$ contains no limit point of K, therefore K must contain at least one limit point of $M_a(Y_1)-K$. But clearly this is impossible, since by Theorem 1 it follows that $S(X_1, Y_1)$ and $S(X_2, Y_2)$ are open subsets of M. Thus the supposition that $S(X_1, Y_1)$ and $S(X_2, Y_2)$ are not mutually exclusive leads to a contradiction. Hence no two of the segments of the collection $[S(X, Y)]$ can have a common point. But since, by Theorem 1, each of these segments is an open subset of M and since $[S(X, Y)]$ is an uncountable collection, this clearly is impossible. Thus the supposition that Q is uncountable leads to a contradiction.

Now let H denote the set of points $F-Q$. Then since F is uncountable and Q is countable, H must be uncountable. By hypothesis, for each point X of H there exists a subcontinuum N_x of M which contains X but which contains no other point whatever of E. Now if X and Y are any two distinct points of H, the continua N_x and N_y can have no point in common. For suppose they do have a point P in common. The point Y belongs either to $M_a(X)$ or to $M_b(X)$, say to $M_b(X)$. Since X and Y belong to H, the segment XY of M must contain at least one point O of F, and O separates A and B in M. But the set of points $M_a(X)+M_b(Y)+N_x+N_y$ is connected and contains both A and B and does not contain the point O. Clearly this is impossible. Thus the supposition that N_x and N_y have a point in common leads to a contradiction. Now since H is uncountable, the collection of

continua $[N_x]$ must be uncountable. But no two of these continua have a common point, and each of them contains at least one cut point of M. This is contradictory to Theorem 5. Thus the supposition that Theorem 6 is not true leads to a contradiction, and the theorem is proved.

3. The order and regularity of the cut points of a continuum

Definitions. The point P of a continuum M is said to be a point of *Menger order n of M** provided that for each positive number ϵ there exists a domain U containing P and of diameter less than ϵ and whose boundary has not more than n points in common with M, and furthermore n is the smallest positive integer such that this property is preserved. A point P of a continuum M will be called a regular or an irregular point of M according as M is or is not connected im kleinen at P.

Theorem 7. *Let G denote the set of all the cut points of any continuum M. Then all save possibly a countable number of the points of G are points of Menger order two of M.*

Suppose, on the contrary, that there exists an uncountable subset H of G no point of which is a point of order two of M. Then by Theorem 3, there exist two points A and B of H which are separated in M by each point of an uncountable subset D of H. It follows by Theorem 6 that there exists at least one point P of D such that every subcontinuum of M which contains P must contain at least one point of D distinct from P. It is easy to see that each of the sets $M_a(P)$ and $M_b(P)$ must be connected and that P is a limit point of a subset D_1 of D belonging to $M_a(P)$ and also of a subset D_2 of D belonging to $M_b(P)$.

Let K_1 denote the continuum $M_a(P)+P$ and K_2 the continuum $M_b(P)+P$. There exists a sequence of points $X_1, X_2, X_3, \cdots$, belonging to D_1 and having P as its sequential limit point and such that for each positive integer n, X_n precedes X_{n+1} on M in the order from A to B. Now let us consider the sequence of intervals $I(X_1, X_2)$, $I(X_2, X_3)$, $I(X_3, X_4), \cdots$, of M. R. G. Lubben has shown† that if N denotes the sequential limiting set of this sequence, then N exists and is closed and connected. Clearly N contains the point P. It must consist of P alone. For suppose it does not. Then N is a subcontinuum of K_1 which contains P. It is readily seen that N cannot contain any point of the set of points

* Cf. K. Menger, *Grundzüge einer Theorie der Kurven*, Mathematische Annalen, vol. 95 (1925), pp. 272–306.

† See an abstract of his paper *Concerning limiting sets* in the Bulletin of the American Mathematical Society, vol. 32 (1926), p. 14. See also Zarankiewicz, loc. cit.

$$L = M_a(X_1) + \sum_{i=1}^{\infty} I(X_i, X_{i+1}).$$

But L is connected and P is a limit point of L. Hence K_2+L is connected and contains both A and B but contains in common with N only the point P. Hence N contains P but contains no other point of D. This contradicts our definition of the point P. Thus it follows that N is identical with P.

Now let ϵ be any positive number. Then since the sequential limiting set of the sequence of intervals $I(X_1, X_2)$, $I(X_2, X_3)$, $\cdots$, consists of just the point P, and since P is the sequential limit point of the sequence of points $X_1, X_2, X_3, \cdots$, it is readily seen that there exists a positive integer k such that if I denotes the set of points $\sum_{i=k-1}^{\infty} I(X_i, X_{i+1})$, every point of I is at a distance less than $\epsilon/4$ from P. Now by Theorem 1, $I(X_k, P)$ is connected, and it contains X_k. It does not contain X_{k-1}, because X_{k-1} precedes X_k on M in the order from A to B. Hence $I(X_k, P)$ must be a subset of $M_b(X_{k-1})$, and since it is a subset also of K_1, it must be a subset of $I+P$. But $I+P$ is of diameter less than $\epsilon/2$. Hence $I(X_k, P)$ is of diameter less than $\epsilon/2$. In an entirely similar manner it is shown that D_2 contains a point Y_k such that $I(P, Y_k)$ is of diameter less than $\epsilon/2$. Then $I(X_k, Y_k)$ contains P and is of diameter less than ϵ. And since the maximal connected subset of $S(X_k, Y_k)$ which contains P must have both X_k and Y_k as limit points, it is easily seen that there exists a domain of diameter less than ϵ which contains P and whose boundary has in common with M just the points X_k and Y_k. Hence P is a point of Menger order two of M. But by supposition P belongs to H, and no point of H is a point of Menger order two of M. Thus the supposition that Theorem 7 is not true leads to a contradiction.

THEOREM 8. *If A and B are any two points of a continuum M, and P is any point belonging to $K(A, B)$ and having the property that every subcontinuum of M which contains P contains at least one point of $K(A, B)$ distinct from P, then P is a point of Menger order two of M.*

The argument given to prove Theorem 7 also proves Theorem 8.

THEOREM 9. *If P is any cut point of a continuum M which is a point of Menger order two of M, then $M-P=M_1(P)+M_2(P)$, where $M_1(P)$ and $M_2(P)$ are connected point sets and P is a point of order one (i.e., an end point in the Menger sense) of each of the continua $M_1(P)+P$ and $M_2(P)+P$.*

THEOREM 10. *If P is any cut point of a bounded continuum M which is a point of Menger order two of M, then P belongs to the boundary of just one complementary domain of M.*

By a theorem of R. L. Moore's*, P belongs to the boundary of at least one complementary domain of M. Suppose, contrary to this theorem, that P belongs to the boundary of two complementary domains R_1 and R_2 of M. Then there exists† a simple closed curve J which encloses P and is of diameter less than half the diameter of R_1 and also less than half the diameter of R_2, and which has in common with M just two points A and B. Then of the two arcs of J from A to B, one of them, say AXB, must belong to R_1, and the other, AYB, must belong to R_2. But from Theorem 9 and a theorem of R. L. Moore's‡ it follows that there exists a simple closed curve C enclosing one of the sets $M_1(P)$ and $M_2(P)$ and not the other (where $M-P=M_1(P)+M_2(P)$, as in Theorem 9), and containing in common with M only the point P. It is easy to see that C must contain at least one point U of AXB and at least one point V on AYB. Then if t denotes the arc of M from U to V which does not contain P, then t contains no point of M. But this is impossible, because U and V belong to different complementary domains of M. Thus the supposition that Theorem 10 is not true leads to a contradiction.

THEOREM 11. *If G denotes the set of all the cut points of a continuum M, then all save possibly a countable number of the points of G are regular points of M.*

Theorem 11 is a direct consequence of Theorem 7 and of Menger's§ theorem that a continuum M is connected im kleinen at each of its points which is a point of finite order of M.

THEOREM 12. *If D denotes the collection of all the cut points $[P]$ of a continuum M such that P is an irregular point of some subcontinuum of M, then D is countable.*

Theorem 12 follows at once from Theorem 7 and the fact that any point P of Menger order two of a continuum M is a regular point of every subcontinuum of M which contains P.

---

* R. L. Moore, *Concerning the common boundary of two domains*, Fundamenta Mathematicae, vol. 6 (1924), pp. 203–213.

† This follows readily from a theorem of R. G. Lubben's. See the abstract of his paper in the Bulletin of the American Mathematical Society, vol. 32 (1926), p. 114.

‡ R. L. Moore, *Concerning the separation of point sets by curves*, Proceedings of the National Academy of Sciences, vol. 11(1925), pp. 469–476, Theorem 2. See also R. G. Lubben, loc. cit.

§ Cf. K. Menger, loc. cit.

4. Miscellaneous results

In this section some theorems will be stated with little or no proof given. These theorems, in general, either follow readily from the above propositions already proved or else they can be proved readily using methods similar to those employed in the above proofs.

Theorem 13. *Let G denote the set of all the cut points of any continuum M. Then G is a subset of the sum of a countable number of bounded subcontinua of M each of which is irreducible between some pair of points.*

Theorem 13 can be proved with the aid of Lemma 2 to Theorem 5 and methods similar to those used in the proof of Theorem 4, together with the theorem of Janiszewski's* that every two points of a bounded continuum M lie together in a subcontinuum of M which is irreducible between those two points.

Theorem 14. *The set G of all the cut points of a continuous curve M (bounded or not) is a subset of the sum of a countable number of simple continuous arcs which belong to M.*

Theorem 15. *If N is any subcontinuum of a continuum M which is irreducible between some pair of points A and B of M, then all save possibly a countable number of the cut points of M which belong to N must separate A and B in M.*

Corollary. *If t is any simple continuous arc belonging to a continuous curve M, and A and B denote the end points of t, then all save possibly a countable number of the cut points of M which lie on t separate A and B in M.*

---

* S. Janiszewski, *Sur les continus irréductibles entre deux points*, Journal de l'Ecole Polytechnique, (2), vol. 16 (1912).

The University of Texas,
Austin, Texas

SOME THEOREMS ON THE CONNECTION BETWEEN IDEALS AND GROUP OF A GALOIS FIELD*

BY

OYSTEIN ORE

Let K be a Galois field of order N and G the corresponding finite group. We shall in this paper study some connections between the ideals in K and the constitution of the group G.

Let $\mathfrak{A}$ be an arbitrary ideal in K. We then introduce a subgroup G_A† of G, which we call *the group of the ideal* $\mathfrak{A}$ and which is defined in the following way. The group G_A consists of all substitutions S in G having the property that for every number α in $\mathfrak{A}$ the conjugate number $\alpha' = S:\alpha$ is also a number in $\mathfrak{A}$.

When $\mathfrak{A} = \mathfrak{P}$ is a prime ideal the group G_P of $\mathfrak{P}$ is equal to the "Zerlegungs" group of $\mathfrak{P}$, studied by Dedekind‡ and Hilbert.§ It is known that the group G_P is a metacyclic group, and that the form of the prime-ideal decomposition of the corresponding rational prime p is closely connected with the properties of this group.

In §1 we study the properties of the general groups G_A and completely determine their construction. In §2 we deal with the properties of the field K_A corresponding to the group G_A, and we prove that K_A is the subfield of K of lowest degree for which there exists an exponent n, such that $\mathfrak{A}^n$ is an ideal in the field. In §3 we study another generalization of the "Zerlegungs" field of Hilbert, namely the fields wherein all numbers are congruent to a rational number (mod $\mathfrak{A}^\alpha$) for an arbitrarily great α, and for a given $\mathfrak{A}$ we determine all the fields having this property.

In another paper I shall use these results in a more elaborate study of the arithmetic of Galois fields and their subfields.

1. **Determination of the group of an ideal.** Let G_A be the group of the ideal $\mathfrak{A}$; then G_A consists of all substitutions in G transforming every number α in $\mathfrak{A}$ into another number $\alpha' = S:\alpha$ also contained in $\mathfrak{A}$. These substitutions obviously form a group, because the product SS' of two substitutions in G_A

* Presented to the Society, October 29, 1927; received by the editors in October, 1927.

† For typographical reasons, italic letters have been substituted, in subscripts and superscripts, for the corresponding German letters.

‡ R. Dedekind, *Zur Theorie der Ideale*, Göttinger Nachrichten, 1894.

§ D. Hilbert, *Grundzüge einer Theorie des Galoisschen Zahlkörpers*, Göttinger Nachrichten, 1894.

must also change α into $\alpha'' = SS' : \alpha$, where α'' is also contained in $\mathfrak{A}$.

When S is a substitution in G_A, then S cannot transform a number β not contained in $\mathfrak{A}$ into a number α' in $\mathfrak{A}$, because from $S:\beta = \alpha'$ it follows that the substitution S^{-1} in G_A must transform α' into β, which is evidently not possible.

When $\mathfrak{A}$ is an ideal in any subfield k of K, the group G_A obviously contains the corresponding subgroup $G^{(k)}$ of G. On the other hand, there exist ideals $\mathfrak{A}$ not contained in k, such that $\mathfrak{A}$ is invariant under all substitutions of the group $G^{(k)}$. In §2 we shall, for a given subfield k, determine all ideals in K having this property.

In particular, when $[a] = \mathfrak{A}$ is a principal ideal and a a rational integer, then G_A is equal to the complete group G. From this remark we can derive a more available criterion for the determination of the substitutions of the group G_A.

The ideal $\mathfrak{A}$ always contains rational numbers, for instance the norm $N = N(\mathfrak{A})$ of the ideal. We can further find a number α in $\mathfrak{A}$ such that the ideals

$$[N],\ [\alpha]/\mathfrak{A}$$

are relatively prime; the principal ideals $[N]$ and $[\alpha]$ consequently have the greatest common ideal factor $\mathfrak{A}$. Let us call α a *primary number* in $\mathfrak{A}$. When therefore β is an arbitrary number in $\mathfrak{A}$, a number γ can always be determined such that

$$\beta \equiv \alpha\gamma \pmod{N}.$$

Consequently if a substitution S transforms α into α', where α' also belongs to $\mathfrak{A}$, the substitution S must be a substitution in G_A, because the arbitrary number β in $\mathfrak{A}$ is transformed into the number

$$\beta' = S:\beta \equiv \alpha'\gamma' \pmod{N},$$

where β' must be a number in $\mathfrak{A}$.

THEOREM 1. *If α is a primary number in an ideal $\mathfrak{A}$, and S a substitution transforming α into α' also belonging to $\mathfrak{A}$, then S is a substitution of the group G_A of $\mathfrak{A}$.*

The group G_A consists of the substitutions of G for which the ideal $\mathfrak{A}$ is invariant. If r_A is the order and $s_A = N/r_A$ is the index of the group G_A, we can write the whole group G as a partition of co-sets corresponding to G_A,

$$G = G_A + G_A V_2 + \cdots + G_A V_{s_A},$$

where V_i are certain substitutions in G. Correspondingly we have s_A different conjugate ideals

$$\mathfrak{A} = \mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_{s_A},$$

where the group of an ideal $\mathfrak{G}_i$ is

$$G_{A_i} = V_i^{-1} G_A V_i.$$

For the norm of $\mathfrak{A}$ in K we therefore have the expression

$$N(\mathfrak{A}) = (\mathfrak{A}_1 \cdots \mathfrak{A}_{s_A})^{r_A}. \tag{1}$$

We shall now examine the construction of the groups G_A and we first mention the following, almost obvious theorem:

THEOREM 2. *The group G_{A^a} of a power is equal to the group G_A of $\mathfrak{A}$.*

We also easily prove

THEOREM 3. *When the norms $N(\mathfrak{A})$ and $N(\mathfrak{B})$ of two ideals $\mathfrak{A}$ and $\mathfrak{B}$ are relatively prime, then the group G_C of the product $\mathfrak{C} = \mathfrak{A}\mathfrak{B}$ is equal to the greatest common subgroup (G_A, G_B) of the groups G_A and G_B.*

When namely S is a substitution in (G_A, G_B), then evidently S belongs to G_C. Conversely, if S is a substitution in G_C, then $\mathfrak{A}$ and $\mathfrak{B}$ must both be invariant with respect to S, because if

$$S{:}\mathfrak{A} = \mathfrak{A}', \quad S{:}\mathfrak{B} = \mathfrak{B}',$$

we have from the definition of G_C

$$\mathfrak{A}'\mathfrak{B}' = \mathfrak{A}\mathfrak{B}. \tag{2}$$

Now the norms $N(\mathfrak{A})$ and $N(\mathfrak{B})$ are relatively prime, so that $\mathfrak{B}$ cannot have any common factor with any of the conjugates $\mathfrak{A}'$ of $\mathfrak{A}$, and consequently we obtain from (2)

$$\mathfrak{A}' = \mathfrak{A}, \ \mathfrak{B}' = \mathfrak{B},$$

and S belongs to (G_A, G_B).

Theorem 3 obviously also holds for an arbitrary number of ideals, whose norms are all relatively prime. The problem of the determination of the general group G_A is therefore reduced to the case where the norm of $\mathfrak{A}$ is the power of a rational prime p, and $\mathfrak{A}$ therefore only contains prime ideals dividing p.

We now prove

THEOREM 4. *If $\mathfrak{A}$ is an ideal containing only prime ideals $\mathfrak{P}$ dividing the same rational prime p, we can uniquely write $\mathfrak{A}$ in the form*

$$\mathfrak{A} = \mathfrak{A}_1^{a_1} \cdots \mathfrak{A}_r^{a_r} \tag{3}$$

where all exponents a_i are different, all $\mathfrak{A}_i$ relatively prime, and in general

$$\mathfrak{A}_i = \mathfrak{P}_i \mathfrak{P}_i' \cdots \qquad (i = 1, 2, \cdots, r)$$

is the product of different, conjugate prime ideals. Then the group G_A is equal to the greatest common subgroup of all groups G_{A_i}.

$$G_A = (G_{A_1}, G_{A_2}, \cdots, G_{A_r}). \tag{4}$$

It is obvious that a unique representation of $\mathfrak{A}$ in the form (3) always exists. If now S is a substitution contained in G', where

$$G' = (G_{\alpha_1}, \cdots, G_{\alpha_r}), \quad \alpha_i = \mathfrak{A}_i^{a_i}, \tag{5}$$

is the greatest common subgroup of all the groups G_{α_i}, then $\mathfrak{A}$ is evidently invariant with respect to S, so that S is contained in G_A.* When on the other hand S is a substitution in G_A, we can prove that S leaves all ideals $\mathfrak{A}_i^{a_i}$ unaltered, and consequently belongs to G'. For if $S: \mathfrak{A} = \mathfrak{A}'$ is the transformed ideal of $\mathfrak{A}$, we can only have $\mathfrak{A} = \mathfrak{A}'$ if at the same time $S: \mathfrak{A}_i^{a_i} = \mathfrak{A}_i^{a_i}$ for all i, because if S transforms a prime ideal $\mathfrak{P}_1$ in $\mathfrak{A}_1$ into, for instance, the prime ideal $\mathfrak{P}_2$ in $\mathfrak{A}_2$, then the ideal $\mathfrak{A}'$ must contain the prime ideal $\mathfrak{P}_2$ to the power $\mathfrak{P}_2^{a_1}$ and consequently $\mathfrak{A} \neq \mathfrak{A}'$, contrary to the definition of S. We therefore have $G_A = G'$ and if we apply Theorem 2 to the expression (5) for G' we obtain the formula (4).

It now only remains to study the groups of ideals having the form

$$\mathfrak{A} = \mathfrak{P}_{s_1} \mathfrak{P}_{s_2} \cdots \mathfrak{P}_{s_t}, \tag{6}$$

where the prime ideals $\mathfrak{P}_s$ are all different, but factors of the same rational prime p.

If G_P is the group of an arbitrary prime ideal dividing p (the "Zerlegungs" group of $\mathfrak{P}$), we obtain from (1)

$$p^f = (\mathfrak{P}_1 \cdots \mathfrak{P}_{s_P})^{r_P}$$

and from this relation we immediately derive the well known result

$$p = (\mathfrak{P}_1 \cdots \mathfrak{P}_{s_P})^e, \quad e = \frac{r_P}{f}, \tag{7}$$

where e and f denote the order and the degree of the prime ideal $\mathfrak{P} = \mathfrak{P}_1$.

The corresponding groups of the prime ideals in (7),

$$G_{P_1}, G_{P_2}, \cdots, G_{P_\pi}, \qquad \pi = s_P,$$

are all conjugate: their greatest common subgroup H is a self-conjugate

* See second foot note on page 610.

subgroup of the complete group G. We write the group G in co-sets corresponding to H:

$$G = H + HV_2 + \cdots + HV_{s_H}, \tag{8}$$

where s_H is the index of H, and

$$V_2, \cdots, V_{s_H}$$

are certain substitutions in G. By means of (8) we construct the factor group G/H of order s_H. The substitutions of H produce no permutation in the order of the prime ideals in (7),

$$\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_{s_P},$$

but all the substitutions of an arbitrary co-set HV_k produce the same permutation

$$\mathfrak{P}_{c_1}, \mathfrak{P}_{c_2}, \cdots, \mathfrak{P}_{c_\pi}, \qquad \pi = s_P,$$

among these ideals. All substitutions

$$C = \begin{pmatrix} 1, 2, \cdots, s_P \\ c_1, c_2, \cdots, c_{s_P} \end{pmatrix} \tag{9}$$

thus obtained form a group simply isomorphic with the factor group, so we can suppose the factor group given in the form of this substitution group. It is well known that to every subgroup of the factor group corresponds uniquely a subgroup of G containing H.

By means of the factor group G/H we can easily determine the group G_A of an ideal $\mathfrak{A}$ having the form (6). The group H is evidently a subgroup of G_A. When, however, S is a substitution belonging to one of the co-sets HV_k in (8), this substitution produces the permutation C (9) among the prime ideals in (7). When therefore the ideal $\mathfrak{A}$ is to be invariant with respect to S, the substitution (9) corresponding to the co-set HV_k must have the form

$$C = C_1 C_2, \tag{10}$$

where C_1 represents an intransitive substitution containing only the indices $s_1, \cdots, s_t$ and similarly C_2 a substitution containing only the remaining indices. The substitutions in the factor group G/H having the form (8) obviously form an intransitive subgroup of G/H and the corresponding subgroup in G must be the group G_A. So we have the following theorem:

THEOREM 5. *Let the self-conjugate group H be the greatest common subgroup of all the groups G_{P_i} corresponding to the different prime ideals $\mathfrak{P}_i$ in p, and*

let G/H denote the factor group of G corresponding to H. Then the group of an ideal

$$\mathfrak{A} = \mathfrak{P}_{s_1} \cdots \mathfrak{P}_{s_t},$$

where all factors are different conjugate prime ideals, is the subgroup of G which corresponds to the intransitive subgroup G_A' of G/H permuting only the indices $s_1, \cdots, s_t$ inter se and correspondingly permuting the remaining indices mutually.

2. **The corresponding fields.** To the group G_A of an arbitrary ideal $\mathfrak{A}$ corresponds uniquely a certain subfield K_A of K. We call K_A *the subfield corresponding to* $\mathfrak{A}$, and we shall now study the arithmetical properties of this field.

We shall first prove a preliminary theorem concerning an arbitrary subfield k of K. Let $G^{(k)}$ be the subgroup of G corresponding to k. As we have already remarked in §1, all the ideals in k are invariant with respect to the substitutions of $G^{(k)}$, but we can also show the existence of ideals in K *not* contained in k having the same property, and our first problem is to determine all these ideals.

Let $\mathfrak{p}$ be a prime ideal in k containing a prime ideal $\mathfrak{P}$ in K as a factor. We can then determine the prime ideal decomposition of $\mathfrak{p}$ in K in the following way: Let G' be the greatest common subgroup of $G^{(k)}$ and the group G_P; then $G^{(k)}$ can be developed in co-sets corresponding to G',

$$G^{(k)} = G' + G'V_2 + \cdots + G'V_s,$$

and it follows easily that $\mathfrak{p}$ is divisible by s different prime ideals

$$\mathfrak{P}_1 = \mathfrak{P}, \mathfrak{P}_2, \cdots, \mathfrak{P}_s.$$

We therefore obtain

$$N_k(\mathfrak{P}) = \mathfrak{p}^{f_k} = (\mathfrak{P}_1 \cdots \mathfrak{P}_s)^r,$$

where N_k indicates the norm in k, f_k is the relative degree of $\mathfrak{P}$, and r the order of the group G'. Consequently we have the prime ideal decomposition*

$$\mathfrak{p} = (\mathfrak{P}_1 \cdots \mathfrak{P}_s)^{e'}, \quad e' = \frac{r}{f_k} \tag{11}$$

where the exponent e' necessarily is a divisor of N, r being the order of a subgroup of G.

If now $\mathfrak{A}$ is an ideal in K, which is invariant with respect to the group $G^{(k)}$, and we suppose $\mathfrak{A}$ to be divisible by exactly $\mathfrak{P}^\lambda$, then $\mathfrak{A}$ obviously must

* See, for example, H. Weber, *Lehrbuch der Algebra*, vol. II, on relative fields.

contain all the prime ideals $\mathfrak{P}_i$ dividing $\mathfrak{p}$ in (11) to the same power $\mathfrak{P}_i^{\lambda}$. Hence we can write

$$\mathfrak{A} = (\mathfrak{P}_1 \cdots \mathfrak{P}_s)^{\lambda} \mathfrak{B},$$

where the ideal $\mathfrak{B}$ is also invariant with respect to $G^{(k)}$, but relatively prime to $\mathfrak{p}$. By reasoning in the same way on $\mathfrak{B}$ we finally conclude that every invariant ideal can be written as a product

$$\mathfrak{A} = \prod_{\mathfrak{p}} (\mathfrak{P}_1 \cdots \mathfrak{P}_s)^{\lambda} \tag{12}$$

corresponding to the different prime ideals $\mathfrak{p}$ in k having common factors with $\mathfrak{A}$. Conversely, if $\mathfrak{A}$ has the form (12), it is evident that the ideal is invariant with respect to $G^{(k)}$.

The necessary and sufficient condition for an invariant ideal can also be expressed in another way. In (11) the exponent e' is a divisor of N. When we therefore form the Nth power of $\mathfrak{A}$, we see from (12) that every term

$$(\mathfrak{P}_1 \cdots \mathfrak{P}_s)^{\lambda N}$$

is equal to a certain power of the corresponding prime ideal $\mathfrak{p}$ in k, so that the ideal $\mathfrak{A}^N$ is an ideal in k.

THEOREM 6. *The necessary and sufficient condition that an ideal $\mathfrak{A}$ be invariant with respect to a group $G^{(k)}$, is that there exist an exponent n such that $\mathfrak{A}^n$ is an ideal in the corresponding subfield k.*

As a corollary we see that when an ideal $\mathfrak{A}$ is invariant relative to all substitutions of the complete group G, then $\mathfrak{A}^N = [a]$ is a rational principal ideal.*

We remark that when $\mathfrak{A}$ is invariant with respect to the substitutions of the group $G^{(k)}$, then $\mathfrak{A}$ is also invariant relative to all subgroups of $G^{(k)}$, and correspondingly it is evident that the ideal $\mathfrak{A}^N$ must be an ideal in all fields having k as a subfield.

Let now K_A be the subfield of K corresponding to the group G_A. Then $\mathfrak{A}$ is evidently an invariant ideal of G_A and by Theorem 6 it therefore follows that an exponent n exists such that $\mathfrak{A}^n$ is an ideal in K_A. But by definition G_A is the greatest subgroup of G for which $\mathfrak{A}$ is invariant, and consequently K_A is the field of lowest degree which contains a power of $\mathfrak{A}$.

THEOREM 7. *The field K_A corresponding to an arbitrary ideal $\mathfrak{A}$ is the subfield of lowest degree for which there exists an exponent n such that $\mathfrak{A}^n$ is an ideal*

* D. Hilbert, *Über die Zerlegung der Ideale eines Zahlenkörpers in Primideale*, Mathematische Annalen, vol. 44 (1894), p. 1.

contained in the field. Every field that contains a power of $\mathfrak{A}$ *contains* K_A *as a subfield.*

As an application we suppose that $\mathfrak{A}=\mathfrak{P}$ is a prime ideal, and hence G_P is the "Zerlegungs" group and K_P the "Zerlegungs" field, introduced by Hilbert. In this case* there exists a prime ideal $\mathfrak{p}_0$ in K_P of degree $f_0=1$ and order $e_0=1$ such that

$$\mathfrak{p}_0 = \mathfrak{P}^s, \quad N_P(\mathfrak{P}) = \mathfrak{p}_0^f, \tag{13}$$

where N_P indicates the relative norm with respect to K_P. From Theorem 7 we conclude that K_P is the field of lowest degree for which an exponent n exists such that $\mathfrak{P}^n$ is an ideal in the field; every other field k having this property contains K_P as a subfield. When n is the least exponent for which $\mathfrak{P}^n$ is an ideal in k, then obviously $\mathfrak{p}=\mathfrak{P}^n$ is a prime ideal in k. In fields not containing K_P as a subfield the prime ideal $\mathfrak{p}$ which is divisible by $\mathfrak{P}$ must also contain some of the conjugates of $\mathfrak{P}$.

THEOREM 8. *The "Zerlegungs" field* K_P *is the field of lowest degree for which there exists an exponent* n *such that* $\mathfrak{P}^n=\mathfrak{p}$ *is a prime ideal in the field. Every other field having this property contains* K_P *as a subfield, and in a field not containing* K_P *the prime ideal* $\mathfrak{p}$ *divisible by* $\mathfrak{P}$ *must also be divisible by other prime ideals.*

3. **The rational congruence-field of an ideal.** The "Zerlegungs" field K_P has also another very important property. As we have already mentioned, the ideal $\mathfrak{p}_0$ in (13) is of first degree and first order. Hence, if ω_P is an arbitrary number in K_P, we have

$$\omega_P \equiv a \pmod{\mathfrak{p}_0},$$

where a is rational, and, because the prime p contains exactly the first power of $\mathfrak{p}_0$, it follows that a congruence

$$\omega_P \equiv a \pmod{\mathfrak{p}_0^\alpha} \tag{14}$$

also holds for all α, where the rational number a obviously depends on α. From (14) and (13) we therefore derive

THEOREM 9. *When* ω_P *is an arbitrary number in the "Zerlegungs" field* K_P, *then a congruence*

$$\omega_P \equiv a \pmod{\mathfrak{P}^\alpha}.$$

holds for all α, *where* a *is a rational number.*

* See, for example, D. Hilbert, *Bericht über die Theorie der algebraischen Zahlen*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 4 (1894–95), §4.

We say that the field K_P is rational (mod $\mathfrak{p}_0^\alpha$), or also K_P is a rational congruence-field (mod $\mathfrak{p}_0^\alpha$). As a generalization of Theorem 9 we could propose to determine all subfields R_A of K, which are rational (mod $\mathfrak{A}^\alpha$), where $\mathfrak{A}$ is an arbitrary ideal in K. We denote by R_A the greatest subfield of K having this property and we see that every other rational field (mod $\mathfrak{A}^\alpha$) is a subfield of R_A. We can also suppose obviously that the ideal $\mathfrak{A}$ is the product of different prime ideals in K.

Before we examine the rational subfields of K (mod $\mathfrak{A}$), we shall give some results on an arbitrary field k, which is rational (mod $\mathfrak{A}_0^\sigma$), where

$$\mathfrak{A}_0 = \mathfrak{p}_0\mathfrak{p}_1 \cdots \mathfrak{p}_r \tag{15}$$

is an ideal in the field. Every number in k then satisfies a congruence

$$\omega \equiv a \pmod{\mathfrak{A}_0^\sigma} \tag{16}$$

and it is obvious that all the prime ideals $\mathfrak{p}_i$ in (15) must be of first degree and first order. But we can also prove that $\mathfrak{A}_0$ cannot contain two different prime ideals, for instance $\mathfrak{p}_0$ and $\mathfrak{p}_1$, dividing the same rational prime p. Because if we form the equation $f(x)=0$ satisfied by a primitive number ω in k, it follows* that $f(x)$ for all α has a decomposition

$$f(x) \equiv (x - a_0)(x - a_1)f_2(x) \pmod{p^\alpha},$$

where $f_2(x)$ is a rational polynomial. When the discriminant of ω_A is exactly divisible by p^δ, we obtain

$$\omega - a_0 \equiv 0 \pmod{\mathfrak{p}_0^{\sigma-\delta}}$$

and at the same time

$$\omega - a_0 \not\equiv 0 \pmod{\mathfrak{p}_1^\delta},$$

so that a congruence (16) cannot hold for all α if $\mathfrak{A}_0$ also contains $\mathfrak{p}_1$. We have therefore proved

THEOREM 10. *If the field k is rational* (mod $\mathfrak{A}_0^\sigma$), *where the ideal $\mathfrak{A}_0$ in k has the form* (15), *then every prime ideal divisor of $\mathfrak{A}_0$ is of first degree and first order, and $\mathfrak{A}_0$ cannot contain more than one prime ideal dividing the same prime p.*

A consequence of this theorem is that every field which is rational with respect to all its ideals is necessarily the rational field R, because every prime

* I refer to my paper *Über den Zusammenhang zwischen den definierenden Gleichungen und der Idealtheorie in algebraischen Körpern* (erste Mitteilung), Mathematische Annalen, vol. 96 (1926), Chap. II, §2.

p must be a prime ideal of the field, and the discriminant is therefore equal to 1, and by a theorem of Minkowski R is the only field having this property.

When $\mathfrak{A}_0$ has the form (15), k must be rational (mod $\mathfrak{p}_i^{a}$) for $i=1, 2, \cdots, r$; but conversely we can conclude, that if k is rational (mod $\mathfrak{p}_i^{a}$) for $i=1, 2, \cdots, r$, then k must be rational (mod $\mathfrak{A}_0^{a}$). When namely $\mathfrak{p}_i$ divides the rational prime p_i, the numbers

$$0, 1, \cdots, p_i^{a}-1$$

form a complete system of incongruent numbers (mod $\mathfrak{p}_i^{a}$), and when we put

$$M_i = \frac{(p_1 \cdots p_r)^{a}}{p_i^{a}}$$

the rational numbers

$$\sum_{i=1}^{r} M_i a_i \qquad (a_i = 0, 1, \cdots, p_i^{a}-1)$$

form a complete system of incongruent numbers (mod $\mathfrak{A}_0^{a}$). It therefore follows conversely, that when an ideal $\mathfrak{A}_0$ satisfies the conditions of Theorem 10, then the field k is necessarily rational (mod $\mathfrak{A}_0^{a}$).

After these general remarks we return to the problem of the determination of all subfields of K that are rational (mod $\mathfrak{A}^{a}$), where $\mathfrak{A}$ is an ideal in K. We first suppose that $\mathfrak{A}$ only contains prime ideals dividing a single prime p, so that $\mathfrak{A}$ has the form

$$\mathfrak{A} = \mathfrak{P}_1 \cdots \mathfrak{P}_t. \tag{17}$$

From Theorem 10 we conclude that in the corresponding field R_A there exists a prime ideal $\mathfrak{p}_0$ of first order and first degree containing $\mathfrak{A}$. The field R_A is obviously a subfield of all the "Zerlegungs" fields of the prime ideals $\mathfrak{P}_i$ in $\mathfrak{A}$, and consequently is also contained in their greatest common subfield. But R_A is generally not equal to this field, as the prime ideals $\mathfrak{P}_i$ can be divisors of two different prime ideals of this field.

We shall now determine the field R_A by means of the corresponding group G'. The group G' must contain the "Zerlegungs" groups G_{P_i} of all divisors $\mathfrak{P}_i$ of $\mathfrak{A}$, and conversely we see that every group G' containing all G_{P_i} corresponds to a field, where the prime ideals $\mathfrak{P}_i$ all divide prime ideals of first order and first degree.

When now in particular G_P is the "Zerlegungs" group of $\mathfrak{P}=\mathfrak{P}_1$ and

$$V_2, V_3, \cdots, V_t$$

a system of substitutions in G such that in general V_i transforms $\mathfrak{P}$ into $\mathfrak{P}_i$, we can prove

THEOREM 11. *The group G' of the field R_A is the least subgroup of G containing the system*

$$G_P + G_P V_2 + \cdots + G_P V_t$$

of co-sets in G.

It is obvious that G' contains all the groups G_{P_i} $(i=1, 2, \cdots, t)$ and in the corresponding field all $\mathfrak{P}_i$ must therefore divide prime ideals of first order and first degree, and this prime ideal $\mathfrak{p}_0$ must be the same for all $\mathfrak{P}_i$, because if $\mathfrak{p}_0$ contains $\mathfrak{P}$ it follows by application of substitutions in G' that $\mathfrak{p}_0$ is divisible by all $\mathfrak{P}_i$ $(i=1, 2, \cdots, t)$.

When, more generally, $\mathfrak{A}$ contains prime ideals dividing different primes, we can write

$$\mathfrak{A} = \mathfrak{A}_1 \mathfrak{A}_2 \cdots,$$

where every factor $\mathfrak{A}_i$ has the form (17). From a former remark it then follows that the corresponding rational field R_A (mod $\mathfrak{A}^a$) is the greatest common subfield of all the fields R_{A_i} $(i=1, 2, \cdots)$.

4. **The inertial group.** I also mention finally that in the same way that we generalized the "Zerlegungs" group of Hilbert we can also obtain an *inertial* group (Trägheitsgruppe) for an arbitrary ideal $\mathfrak{A}$. The inertial group T_A consists of all substitutions T having the property that

$$T:\omega \equiv \omega \pmod{\mathfrak{A}}$$

for every number ω in K.

It is obvious that the group T_A so defined is a subgroup of G_A. The construction of the general group T_A is simple. We obtain the following from the definition of T_A: The inertial group of a composite ideal $\mathfrak{A}$ is equal to the greatest common subgroup of the different prime ideal powers $\mathfrak{P}^a$ dividing $\mathfrak{A}$.

When $\mathfrak{A}=\mathfrak{P}$ is a prime ideal, the group T_P is the "Trägheits" group of Hilbert. When $\mathfrak{A}=\mathfrak{P}^a$, $a>1$, Hilbert introduced the notation "Verzweigungs" group (einmal, zweimal etc. überstrichene). It may perhaps be more natural and simpler to use the general conception of the inertial groups and so totally avoid the term "Verzweigungsgruppe."

There are obviously only a finite number of ideals for which T_A contains substitutions different from the identical substitution.

YALE UNIVERSITY,
NEW HAVEN, CONN.